

How People Use ChatGPT: Conversation-Level Evidence from India, Nigeria, Brazil and Pakistan

SHREYASI ROY CHOWDHURY, Independent researcher, India

KIRAN GARIMELLA, Rutgers University, USA

Recent reports from OpenAI and Anthropic have begun to characterize how hundreds of millions of people use conversational AI, but these analyses rely on aggregated, privacy-preserving indicators with fixed taxonomies and inferred demographics that cannot be re-analyzed by outside researchers. We provide a complementary, conversation-level view: complete ChatGPT exports comprising 202,590 conversations from 1,252 users across India, Nigeria, Brazil, and Pakistan, paired with self-reported age and gender and spanning over three years of use (December 2022–February 2026). Using the same 24-category taxonomy as Chatterji et al. [12], we find that our sample over-indexes on information-seeking and writing and under-indexes on technical help relative to their reported global averages, with substantial heterogeneity across the four countries. Unsupervised topic discovery reveals use cases that a fixed taxonomy cannot see: health and wellness emerges as a highly prevalent theme in India (10%) and Brazil (17%) and a top-5 theme in Pakistan, and finance and online-earning strategies account for 10–17% of conversations, alongside translation needs, religious queries, and emotional self-reflection that vary sharply by country. Classifying conversations by task purpose (work, coursework, or personal) shows that the majority of usage is personal (55–64%), with health advice, translation, and everyday problem-solving far outweighing workplace applications. Usage patterns vary markedly by gender and age: for instance, women in India, Pakistan, and Nigeria use ChatGPT for coursework at substantially higher rates than men (up to 33% vs. 20%), and younger users concentrate on education and programming while older users shift toward finance, career development, and civic topics. Over time, users have shifted from primarily seeking information to increasingly delegating tasks to the model, and reflective, companion-like conversations are growing slowly but consistently across all four countries. These findings suggest that evaluating AI’s impact through workplace productivity alone misses the dominant mode of use in these markets: ChatGPT functions less as a professional tool than as everyday infrastructure for health information, language access, education, and economic navigation—services whose value is real but largely invisible to standard productivity metrics.¹

ACM Reference Format:

Shreyasi Roy Chowdhury and Kiran Garimella. 2018. How People Use ChatGPT: Conversation-Level Evidence from India, Nigeria, Brazil and Pakistan. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 50 pages. <https://doi.org/XXXXXXX.XXXXXXX>

1 Introduction

ChatGPT is now one of the most rapidly adopted technologies in history, with over 700 million monthly active users as of early 2025 [12]. As conversational AI becomes embedded in everyday life at this scale, a fundamental empirical

¹Data and code: <https://github.com/gvrkiran/how-people-use-chatgpt-data>

Authors’ Contact Information: Shreyasi Roy Chowdhury, shreyasi.rc12@gmail.com, Independent researcher, Kolkata, India; Kiran Garimella, kiran.garimella@rutgers.edu, Rutgers University, New Brunswick, NJ, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

Manuscript submitted to ACM

question has moved to the centre of AI research, economics, and public policy: what are all these people actually doing with it?

The answer so far comes from a small number of sources, each with significant limitations. OpenAI’s own analysis of roughly 700 million weekly ChatGPT conversations [12] and its India-specific Signals report [28] are drawn from server-side logs at enormous scale; Anthropic’s Economic Index and its geographic follow-ups [3, 4] do the same for Claude. These platform reports are invaluable for establishing broad trends, but they are, by design, aggregated and privacy-preserving: they classify conversations into a fixed taxonomy chosen in advance, infer demographics from user names or IP addresses, and release only summary statistics. An outside researcher cannot re-run a different classifier, look at the raw conversations, extract tool-use metadata, or test a hypothesis the platform operators did not report on. The WildChat corpus [42] is conversation-level and publicly available, but its users self-selected into a free research proxy, which strongly skews the sample toward technically sophisticated users; our own cross-dataset comparison (Appendix C) shows that WildChat overestimates programming by 4–7× relative to a broader sample. Recent data-donation studies [19, 24] have begun to collect naturalistic ChatGPT conversation histories at the individual level, demonstrating the feasibility of this approach, but focus on European and U.S. populations and do not provide cross-country comparisons or demographically grounded samples from the countries where ChatGPT’s user base is growing fastest.

What this paper contributes. We present a conversation-level analysis of ChatGPT usage across four countries (India, Nigeria, Brazil, and Pakistan) based on complete conversation exports from 1,252 recruited users, paired with self-reported age and gender and spanning over three years of use (December 2022 through February 2026). To our knowledge, this is the first study to combine full, multi-turn ChatGPT conversation logs with verified demographic metadata at this scale for any of these countries, and the first to enable direct cross-country comparison on a common analytical pipeline.

We deliberately apply multiple classification methods to the same corpus, because each reveals something different. First, we replicate the 24-category taxonomy of Chatterji et al. [12] at the conversation level, which makes our results directly comparable to OpenAI’s global averages and lets us measure how these four countries deviate from worldwide patterns. Second, we run unsupervised topic discovery (BERTopic [21]), which helps us find fine-grained topics in the data rather than projecting onto categories designed elsewhere. This turns out to matter: the unsupervised pipeline surfaces culturally specific use cases, such as health and wellness, religious queries, translation needs, digital-entrepreneurship strategies, that the predefined taxonomy absorbs into generic buckets like “Seeking Information.” Third, we replicate two classifiers from Anthropic’s Economic Index [3]: one that distinguishes whether users are seeking information, delegating tasks, or engaging in reflective and emotional conversation; and one that labels each conversation as work-related, coursework-related, or personal. Together, these three lenses answer not only *what* users talk about, but *how* they engage with the model and *why* (for work, school, or personal life). Finally, we extract raw metadata directly from the conversation exports, such as model versions, tool invocations, prompt and response lengths, conversation depth, subscription tier, none of which is available in any published platform report.

Main findings. We organize the results around five analytical lenses, each applied to the same corpus:

- (1) **Topic distribution (Section 5.1).** Under the OpenAI taxonomy, users in our four-country sample over-index on *Seeking Information* (+4.0pp) and *Writing* (+2.5pp) and under-index on *Technical Help* (−5.5pp) relative to global averages [12], with substantial per-country heterogeneity. Gender-conditional patterns (men toward technical

105 help and information seeking, women toward writing, practical guidance, and self-expression) mirror, at the
106 individual-conversation level, the name-inferred patterns reported by OpenAI for India [28].

- 107 (2) **Unsupervised topic discovery (Section 5.2).** The unsupervised pipeline surfaces use cases that the predefined
108 taxonomy cannot see: *health and wellness* as a highly prevalent theme in India (10.1%) and Brazil (17.3%) and a
109 top-5 theme in Pakistan; *religious queries* in Nigeria and Pakistan; *Urdu-English translation* as the top cluster in
110 Pakistan; *self-reflection and emotional support* as a top-5 cluster in Brazil; and *online earning* strategies in India
111 and Pakistan. These culturally specific clusters can only be found with conversation-level access to the data and
112 an open-vocabulary method.
- 113 (3) **Task purpose (Section 5.3).** Personal use dominates in every country (55–64% of conversations), and coursework
114 is comparable in prevalence to work (19.6% vs. 20.8% pooled). Writing and technical help make up 45–50% of work
115 conversations, but the majority of all usage is non-work: health, education, translation, everyday problem-solving.
116 Female users engage with coursework at substantially higher rates than male users in India (26% vs. 18%), Pakistan
117 (33% vs. 20%), and Nigeria (24% vs. 20%).
- 118 (4) **Intent evolution (Section 5.4).** Users have shifted from primarily seeking information to increasingly delegating
119 tasks to the model: task-delegation conversations rose from below 20% in early 2023 to 30–35% by late 2025.
120 Meanwhile, reflective, emotional, and companion-like conversations, though still a small share, drift slowly but
121 consistently upward across all four countries, a trend worth tracking as it carries implications for parasocial
122 attachment and the role of AI in emotional support.
- 123 (5) **Raw metadata (Section 5.5).** Model adoption follows platform defaults rather than active user selection;
124 advanced features (web search, code interpreter, image generation) are used in fewer than 16% of conversations;
125 and ChatGPT Plus subscribers are concentrated almost entirely in India, with near-zero adoption in Nigeria and
126 Pakistan, where the \$20/month subscription represents 10–20% of median monthly income. Language detection
127 reveals that personal conversations are 10–17 percentage points less likely to be in English (or Portuguese in
128 Brazil) than work conversations, and that *Expressing* conversations are conducted in the user’s native language
129 at substantially higher rates than *Doing* conversations, suggesting that users revert to the language they think in
130 for emotional and reflective exchanges.

131 We additionally compare our findings against the WildChat corpus [42], a publicly available dataset of ChatGPT
132 conversations collected via a research proxy (Appendix C). WildChat over-represents programming by 4–7× relative
133 to our recruited sample and under-represents the information-seeking and practical-guidance queries that dominate
134 broader usage, illustrating how collection methodology shapes the empirical picture of AI use.

135 Taken together, these results paint a picture of ChatGPT usage that challenges the dominant framing of conversational
136 AI as primarily a workplace productivity tool. The majority of usage is personal, and the most prevalent topics are
137 health, education, translation, and everyday practical guidance. At the same time, usage is far from uniform: what
138 people do with ChatGPT varies substantially by country, gender, and age, in ways that reflect local labour markets,
139 linguistic needs, educational systems, and cultural context. Women’s disproportionate use of ChatGPT for coursework
140 in India, Pakistan and Nigeria, the concentration of religious queries in Pakistan, and the prevalence of translation
141 in multilingual societies are signals that the value and risks of conversational AI will be distributed unevenly across
142 populations. Understanding these patterns, at the conversation level and with demographic grounding, is a prerequisite
143 for informed policy. This study is a first step. Data donation, in which users export and share their conversation histories
144
145
146
147
148
149
150
151
152
153
154
155
156

with informed consent, is a scalable and replicable methodology for building the empirical base that policymakers, researchers, and platform designers need. We develop these implications in Section 6.

Paper structure. Section 3 describes the dataset and the WildChat corpus used for cross-validation. Section 4 details the classification pipelines. Section 5 presents results organized around the five analytical lenses. Section 6 discusses implications for economics, policy, and product design. All code, classifier prompts, and aggregate tables are released for replication.

2 Related Work

Our work sits at the intersection of five research threads: large-scale studies of LLM usage, AI adoption in the Global South, gender and age disparities in technology adoption, methodological work on classifying user–AI interactions, and the emerging literature on affective and companion-like use of conversational AI.

2.1 Large-Scale Studies of LLM Usage

The earliest efforts to characterize real-world LLM usage relied on research-proxy corpora. WildChat [42] released approximately one million ChatGPT conversations (expanded to 4.8M) collected through a free Gradio frontend; LMSYS-Chat-1M [43] did the same through a model-comparison platform. Both have the advantage of being conversation-level and open, but their collection methods skew the sample toward technically sophisticated users comfortable with non-standard interfaces.

The second wave of evidence has come from the platforms themselves. Chatterji et al. [12] analyzed approximately 700 million weekly ChatGPT conversations over a 15-month period, classifying each into a 24-category taxonomy over seven coarse domains (Practical Guidance, Seeking Information, Writing, Technical Help, Self-Expression, Multimedia, Other) and a three-way *Asking/Doing/Expressing* intent classification. They reported that only 27% of conversations are work-related and that self-expression is the fastest-growing category. OpenAI’s Signals initiative [28] has since released a country-specific companion report for India, providing aggregated age, gender (inferred from names), and topic breakdowns. Anthropic’s Economic Index [3] and its geographic follow-up [4] perform analogous analyses for Claude.ai, adopting the same *Asking/Doing/Expressing* frame and introducing additional measures: a work/coursework/personal task-purpose label, multitasking indicators, and educational-requirements scoring.

A third, emerging line of work collects conversation-level data directly from users through data-export or data-donation mechanisms. Boeschoten et al. [10] introduce a framework for privacy-preserving collection of digital trace data via data donation, treating consent-based donation as a first-class research method. Building on this approach, Karnam et al. [24] gathered 825K ChatGPT interactions from 300 users via GDPR data-export rights and documented a shift from functional to socially framed usage, including substantial growth in health and mental-health queries and rising anthropomorphization of the model. Fang et al. [19] collected 48,495 conversations from 82 U.S.-based adults through a privacy-preserving “wrapped”-style pipeline, finding a mix of instrumental and reflective use and noting that heavier users engage in proportionally more reflective exchanges. Deng et al. [16] analyze a smaller corpus of publicly shared ChatGPT conversations and provide an early user portrait, but face selection-bias concerns since users self-selected into sharing. Tamkin et al. [35] introduce Clio, a privacy-preserving bottom-up topic-discovery system applied to millions of Claude conversations; this is closest in spirit to our unsupervised pipeline, but the underlying conversations are not released and the outputs are aggregate. Together these studies demonstrate the feasibility and value

209 of user-donated conversation data, though none covers the Global South countries or the cross-country comparative
210 framing that motivates our work.

211 These platform reports are invaluable for benchmarking, but three limitations motivate complementary work. First,
212 they are aggregated, privacy-preserving summaries; outside researchers cannot re-run a classifier, extract tool-use
213 metadata, or test a hypothesis the companies did not report on. Second, demographic information is inferred from
214 names or IP addresses rather than self-reported. Third, the topic taxonomy is fixed in advance and cannot be re-opened
215 to discover culturally specific usage patterns. Our study directly addresses these limitations by working with complete
216 conversation exports and self-reported demographics from a recruited cohort across four Global South countries, while
217 using the *same* classifiers the platform reports use so that the two bodies of evidence are directly comparable.
218
219
220

221 2.2 AI Adoption in the Global South

222 A large literature documents persistent digital divides in the Global South along dimensions of infrastructure, literacy,
223 language, and economic access [1, 38, 39]. Mobile-first access patterns [17], linguistic under-representation [23], and
224 data-quality challenges [33] have all been shown to shape how users in these markets engage with digital technologies.
225 For AI specifically, Arora and Rathi [6] examined how AI tools are perceived in Indian workplaces, and Chinchure
226 et al. [13] analyzed early demographic patterns in AI chatbot usage. The platform reports [4, 28] are now beginning
227 to characterize AI usage in specific Global South markets, but they do so only at the aggregate level for individual
228 countries. Ours is, to our knowledge, the first conversation-level cross-country comparison on a common classification
229 pipeline for India, Nigeria, Brazil, and Pakistan.
230
231
232

233 2.3 Gender, Age, and Technology Adoption

234 Gender disparities in technology access and usage have been extensively documented [2, 5, 22]. In the Global South,
235 gender gaps in internet and smartphone access remain substantial, shaped by cultural norms, economic constraints,
236 and educational disparities [20]. For LLMs specifically, Chatterji et al. [12] estimated, using name-based inference, that
237 women’s share of ChatGPT users has grown from 35% to 48% globally, and OpenAI [28] provides a similar decomposition
238 for India. Because name-based inference is known to misclassify gender-ambiguous names (an especially relevant
239 concern for Indian names), our self-reported gender labels provide a useful validation check. We emphasize, however,
240 that we interpret gender-conditional *usage patterns* rather than gender *composition* as findings, since the composition
241 of our recruited cohort reflects the Clickworker sampling frame.
242
243
244
245
246

247 2.4 AI in Education and Work

248 A large literature focuses on the workplace impact of generative AI. Eloundou et al. [18] estimated that 80% of the U.S.
249 workforce could see at least 10% of their tasks affected by LLMs. Field experiments document substantial productivity
250 gains: Brynjolfsson et al. [11] study 5,179 customer-support agents and find heterogeneous improvements concentrated
251 among less-experienced workers; Noy and Zhang [27] report a 40% reduction in completion time and an 18% quality
252 improvement on professional writing tasks; Dell’Acqua et al. [14] document a “jagged” capability frontier for knowledge
253 workers at Boston Consulting Group, with large gains inside the frontier and quality losses outside it. Bick et al. [9]
254 use nationally representative U.S. surveys to chart rapid generative-AI adoption both at work and at home, though
255 they do not observe conversation content. On the platform side, Anthropic [3] report that coding, writing, and analysis
256 dominate professional Claude usage. The geographic Claude analysis [4] finds that software and engineering tasks
257
258
259
260

261 account for the largest share of work-related Claude usage in India, reflecting India’s IT services workforce, and OpenAI
262 [28] finds that work-related ChatGPT use in India is concentrated in writing and technical help.

263 In education, AI chatbots have attracted significant attention [7, 15, 25, 41], including as a partial substitute for
264 tutoring or reference material where these are scarce [40]. Ravšelj et al. [32] report a global student survey ($n = 23,218$ in
265 109 countries) describing common uses such as brainstorming, summarizing, and locating research. Our cross-country
266 view on a common pipeline complements these India-focused aggregates and survey results with conversation-level
267 evidence from Brazil, Nigeria, and Pakistan, and shifts the empirical focus from work to everyday assistance: personal
268 use is the majority in our four-country sample, and even within work the dominant tasks are writing, drafting, and
269 job-search support rather than coding.
270
271

272 2.5 Culturally Situated AI Use

273 A separate strand of research argues that LLM behavior and the value users extract from it are culturally situated
274 rather than universal. Tao et al. [36] show that several frontier LLMs are biased toward English-speaking and Protestant
275 European cultural values, and that cultural alignment can be partially recovered via prompting. Naous et al. [26]
276 demonstrate cultural-context failures in multilingual and Arabic monolingual models in scenarios involving religion,
277 food, and naming. On the user side, aggregate platform reports such as [28] and [4] provide country-level decompositions
278 but classify against fixed, Western-anchored taxonomies. Our unsupervised pipeline surfaces use cases (religious queries,
279 Urdu–English translation, online-earning strategies, self-reflection in Portuguese) that such taxonomies have no category
280 for, and our language analysis shows that users systematically revert to their native language when conversations turn
281 reflective or affective.
282
283
284
285
286

287 2.6 Affective and Companion-Like Use of Conversational AI

288 A growing literature studies the emotional and relational dimension of human–LLM interaction. Phang et al. [31]
289 combine platform-scale affective-cue analysis on ChatGPT with a controlled survey and randomised trial, and find
290 that affective use is concentrated among a small subset of users but has measurable well-being implications. Earlier
291 human–chatbot research, especially Skjuve et al. [34]’s study of Replika users, established that some users develop
292 genuinely social relationships with conversational systems, anticipating the parasocial concerns raised by Turkle [37].
293 Our *Expressing* analysis (Section 5.4) connects to this literature in two ways: a qualitative reading of approximately 100
294 *Expressing* conversations per country shows that what platform classifiers flag as emotional use is rarely pure affect
295 but a hybrid mode that fuses self-disclosure with an instrumental ask, and a language signal (native-language use is
296 highest in *Expressing* and lowest in *Doing*) suggests that users treat ChatGPT differently depending on whether they
297 are delegating a task or processing a feeling.
298
299
300
301
302

303 2.7 Classifying User–AI Interactions

304 Methods for classifying the topics and intents of user–AI interactions range from rule-based taxonomies [30] to
305 unsupervised clustering [43] to LLM-based classification [12]. BERTopic [21] has emerged as a widely adopted method
306 for discovering topic structure in conversational data, combining transformer embeddings with hierarchical clustering.
307 We apply three classifiers in parallel (the OpenAI 24-category taxonomy, BERTopic over Gemini embeddings, and
308 Anthropic’s task-purpose and *Asking/Doing/Expressing* classifiers) so that our findings are both directly comparable to
309 the platform reports and capable of surfacing patterns that a fixed taxonomy cannot see.
310
311

3 Dataset

3.1 Data Collection

We recruited participants from four Global South countries (India, Nigeria, Brazil, and Pakistan) through Clickworker, a global crowdsourcing platform. Participants were required to be active ChatGPT users and were asked to export their complete ChatGPT conversation history using OpenAI’s built-in data export feature, which generates a comprehensive archive including all conversations, model metadata, and account information. Participants also provided basic demographic information (age and gender) during registration. We created a simple interface for the participants to export and upload their ChatGPT histories; after the user downloads their export from OpenAI and before it is uploaded to our servers, a client-side script runs locally on the participant’s machine to strip personal information (names, email addresses, etc.). Participants were compensated for their participation. The data collection happened between December 2025–Feb 2026. Our approach parallels concurrent work by Karnam et al. [24], who collected ChatGPT exports via GDPR data-export rights, and by Fang et al. [19], who designed a privacy-preserving “wrapped”-style pipeline for the same purpose; the key difference is our focus on recruited, demographically profiled users across four Global South countries rather than convenience samples in Europe or the United States.

Table 1 summarizes the key characteristics of our dataset. In total, we collected conversation exports from 1,252 users encompassing 202,590 unique conversations, spanning from December 2022 (shortly after ChatGPT’s launch on November 30, 2022) through February 2026, a period of over three years covering the introduction of GPT-3.5, GPT-4o, and the GPT-5 family.

Table 1. Dataset overview by country. The last four columns describe the distribution of conversations per user.

Country	Users	Conversations	Conversations per User			Date Range
			Mean	Median	P25–P75	
India	557	88,958	159.7	51	11–197	Dec 2022 – Feb 2026
Nigeria	243	44,114	181.5	91	20–225	Dec 2022 – Dec 2025
Brazil	246	40,067	162.9	62	26–159	Dec 2022 – Dec 2025
Pakistan	206	29,451	143.0	55	13–153	Dec 2022 – Dec 2025
Total	1,252	202,590	161.8	60	15–193	Dec 2022 – Feb 2026

India contributes the largest share of users (44.5%) and conversations (43.9%), followed by Nigeria (19.4% of users), Brazil (19.6%), and Pakistan (16.5%). The high variance in conversations per user (IQR of 15–193 overall) reflects a mix of casual and power users across all countries, with Nigeria having the highest median engagement (91 conversations per user). Monthly active users and conversation volume by country are shown in Figure 37.

Table 2 presents the demographic composition of our sample. Overall, 65.9% of users are male and 34.1% are female. However, this gender distribution varies substantially across countries: India (77.2% male) and Pakistan (75.7% male) show the most skewed ratios, while Brazil is the only country where female users outnumber males (55.3% female). Nigeria shows near-parity with a slight male majority (53.1%).

Age distributions also differ markedly: India and Pakistan skew young (median age 25 and 24, respectively, with over 50% in the 18–25 bracket), while Brazil is notably older (median 32, with 37% over 36). Nigeria falls in between (median

Table 2. Demographic composition by country, gender, and age group (user counts).

Country	Gender		Age Group			Mean Age
	Male	Female	18–25	26–35	36+	
India	430 (77.2%)	127 (22.8%)	287 (51.5%)	187 (33.6%)	83 (14.9%)	27.8
Nigeria	129 (53.1%)	114 (46.9%)	56 (23.0%)	133 (54.7%)	54 (22.2%)	31.0
Brazil	110 (44.7%)	136 (55.3%)	45 (18.3%)	110 (44.7%)	91 (37.0%)	33.3
Pakistan	156 (75.7%)	50 (24.3%)	118 (57.3%)	55 (26.7%)	33 (16.0%)	27.1
Total	825 (65.9%)	427 (34.1%)	506 (40.4%)	485 (38.7%)	261 (20.8%)	29.4

30). These demographic differences likely reflect both the age profiles of Clickworker participants in each country and broader patterns of technology adoption. Figure 1 visualizes the demographic composition.

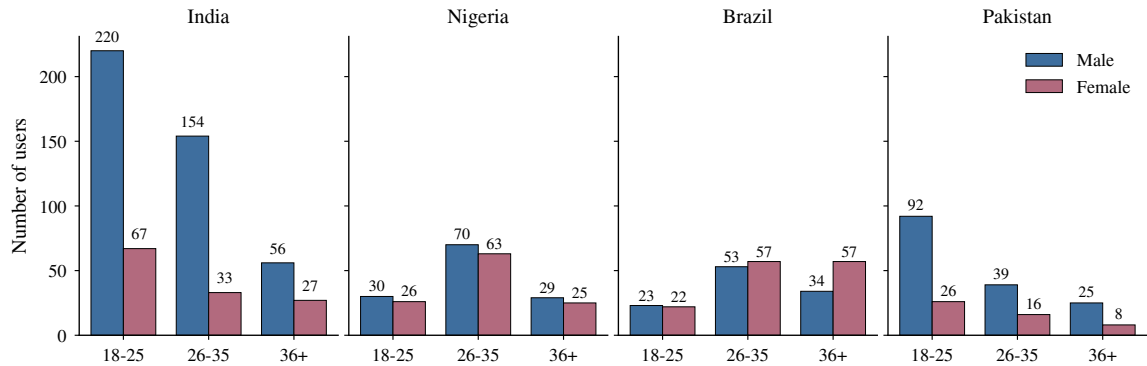


Fig. 1. Demographic composition by country, gender, and age group. Bars show the share of each country’s participants falling in each age×gender cell (so panels are directly comparable despite different per-country sample sizes); labels on bars are within-country percentages.

Nomenclature and fields. Each participant’s archive contains a complete set of their ChatGPT sessions, and their account-level metadata. We adopt the nomenclature implicit in that structure: a *conversation* is a single ChatGPT session (one page in the ChatGPT web or mobile UI) and contains a back-and-forth sequence of *messages*. Each message carries the sender role (user, assistant, or tool), the text content, and per-message metadata including fields identifying the underlying model version and records of any tool invocations (web search, code interpreter, image generation, memory updates, file uploads). From `user.json` we obtain each participant’s subscription tier (Free, Plus, or Pro) and self-reported age and gender.

Note on sample bias. The conversations analyzed in this paper are a small fraction of the ChatGPT user base in each country, and the sample is *not* designed to be demographically representative: it is a convenience sample of Clickworker participants willing to export and share their complete ChatGPT history for compensation, and no stratified effort was made to match the broader national distributions of age, gender, education, or urban/rural residence. Results should therefore be interpreted as patterns *conditional on* being an active, digitally literate, Clickworker-accessible ChatGPT user,

not as estimates of country-wide prevalence. Despite these limits, complete, user-level ChatGPT exports paired with self-reported demographics are, to our knowledge, not publicly available at this scale for any of these four countries, and the dataset offers a first conversation-level view of real-world usage that complements the aggregated server-side indicators that OpenAI and Anthropic have released. We release the anonymized per-conversation classification labels, demographic metadata, and unsupervised cluster assignments at <https://github.com/gvrkiran/how-people-use-chatgpt-data>.

3.2 WildChat Dataset

Ours is not the first publicly analyzable corpus of in-the-wild ChatGPT conversations. The best-known alternative is WildChat [42], which we use in Appendix C as a cross-dataset validation source. We briefly describe it here for completeness.

WildChat was constructed by offering free access to ChatGPT via a Gradio-based web interface in exchange for users' affirmative, consensual opt-in to share their conversation transcripts anonymously for research.

The original WildChat release contains over one million timestamped conversations comprising more than 2.5 million interaction turns, collected between April 2023 and May 2024 via APIs backed by GPT-3.5-Turbo and GPT-4 [42]. An expanded version, WildChat-4.8M, extends the corpus to approximately 4.8 million conversations. Roughly 24-26% of conversations in the original release used the GPT-4 API, with the remainder using GPT-3.5-Turbo. Each conversation record contains the full conversation history, timestamp, the model version used, toxicity flags from the OpenAI Moderation API, hashed IP addresses, and coarse geographic metadata (state and country). WildChat covers more than 60 languages, offering a linguistically diverse outlook of global ChatGPT usage. A key methodological limitation of WildChat is that, users who interact with a dedicated research platform are likely to be limited by the observer effect, participants who know their data is being collected may alter their behaviour. Additionally, its collection method, although consent-based, tends to skew toward users seeking free access to ChatGPT.

Because WildChat is freely available, it has become a de facto reference corpus for empirical studies of ChatGPT usage, and findings drawn from it are sometimes treated as generalizable. Our recruited cohort gives us a natural point of comparison: we apply the same topic classifier to both corpora and, in Appendix C, quantify how the proxy-based sampling frame systematically over-represents technical and programming-oriented use at the expense of the information-seeking and practical-guidance queries that dominate broader user populations.

4 Methods

This section describes the classification pipelines applied to the conversation corpus and a few conventions that are shared across all subsequent analyses. At a high level, every conversation is passed through three independent classifiers: (i) OpenAI's published 24-category topic taxonomy, (ii) unsupervised topic discovery via BERTopic with Gemini embeddings, and (iii) a replication of two of the Anthropic Economic Index's classifiers (*Asking/Doing/Expressing* intent and *work/coursework/personal* task purpose). The first two answer "what is this conversation about?" from two complementary angles (a fixed, directly comparable taxonomy and an open-vocabulary clustering); the third answers "what kind of task is the user doing?" Raw metadata (model version, tool invocations, subscription tier, conversation depth, prompt/response lengths) is extracted directly from the JSON export described in Section 3. All timestamps are stored as UTC Unix timestamps and converted to the primary timezone of each country for local time-of-day analyses: IST (UTC+5:30) for India, WAT (UTC+1) for Nigeria, BRT (UTC-3) for Brazil, PKT (UTC+5) for Pakistan; trend analyses aggregate at monthly granularity and exclude months with fewer than 20 conversations.

4.1 Topic Classification

We classify conversation topics using two complementary methods.

4.1.1 OpenAI Taxonomy. To enable direct comparison with Chatterji et al. [12], we apply their published 24-category fine-grained taxonomy, organized into 7 coarse domains: Practical Guidance, Seeking Information, Writing, Technical Help, Self-Expression, Multimedia, and Other/Unknown. We use the GPT-4o API with the same classification prompt as Chatterji et al., localized only where necessary (e.g., category name consistency).

4.1.2 BERTopic. The OpenAI taxonomy is designed to be stable across hundreds of millions of conversations and, for that reason, is deliberately coarse: a bucket like *Seeking Information* covers anything from symptom interpretation to celebrity trivia to legal clarifications. A privacy-preserving, server-side methodology cannot easily look inside these buckets, but a conversation-level corpus can. We therefore complement the supervised classifier with an unsupervised pipeline that is free to discover whatever structure exists in the data.

Concretely, we embed each conversation using Google’s `gemini-embedding-001` model (3,072-dimensional vectors computed over the first ten user–assistant turns), run MiniBatch *K*-means with $k = 500$ to obtain semantically narrow micro-clusters, and then merge the 500 centroids via agglomerative hierarchical clustering with cosine distance until roughly 50 interpretable top-level topics remain per country. Cluster labels are drafted by `gpt-4o-mini`, adjudicated by Claude Sonnet 4.6 acting as a judge (low-confidence or non-homogeneous clusters are flagged for relabelling or reassignment), and manually reviewed by the authors. The final pipeline yields 50, 45, 36, and 53 topics for India, Nigeria, Brazil, and Pakistan respectively, covering 91–95% of conversations per country; remaining low-confidence conversations are reassigned to existing clusters by embedding-based cosine similarity with a per-cluster *z*-score threshold of 1.5. Complete hyperparameters, dendrograms, and the full judge prompts are documented in Appendix A.

4.2 Intent and Task-Purpose Classifiers

Chatterji et al. [12] introduced a set of conversation-level classifiers that measure *how* a user is using the model, orthogonally to the topic of the conversation; Anthropic [3] subsequently adopted the same framework for Claude. We replicate two of these classifiers on our corpus using the GPT-4o API with the prompts released with the Anthropic Economic Index, applied independently to each conversation. Concretely:

- **Asking/Doing/Expressing.** The three-way intent classification introduced by Chatterji et al. [12]: *Asking* conversations seek information or decision support, *Doing* conversations delegate a task (drafting, transforming, coding, executing) to the model, and *Expressing* conversations use the model for reflection or emotional communication.
- **Work / coursework / personal.** Whether the conversation is work-related, coursework-related, or personal.

4.3 Language Detection

We detect the language of each conversation using the `langdetect` library (with a fixed seed for reproducibility) applied to the concatenated user messages, using the first 5,000 characters; conversations with fewer than 20 characters of user text are marked as *too short* for detection. This yields a per-conversation language code and confidence score, which we use to analyse language choice as a function of task purpose and conversational intent (Section 5.5).

5 Results

We organize our results around five analytical lenses applied to the same 1,252-user, 202,590-conversation corpus: (1) the OpenAI 24-category taxonomy [12] applied at the conversation level; (2) unsupervised topic discovery, which surfaces culturally specific themes that a predefined taxonomy absorbs into generic buckets; (3) the *Asking/Doing/Expressing* intent classifier of Chatterji et al. [12]; (4) the Anthropic-style work/coursework/personal task-purpose classifier [3]; and (5) raw-metadata descriptives on model adoption, tool use, conversation depth, prompt/response lengths, and subscription tier that are simply not available in any published platform report. Each lens is examined at the aggregate, per-country, gender-conditional, age-conditional, and temporal levels. A cross-dataset comparison against the WildChat corpus [42] is reported in Appendix C.

5.1 OpenAI Taxonomy: Topic Distribution and Demographic Conditioning

We apply the OpenAI 24-category fine-grained taxonomy (organized into seven coarse domains) from Chatterji et al. [12] at the conversation level using GPT-4o; details of the classifier are in Section 4. An aggregate treemap of the topic distribution appears in Figure 29.

5.1.1 Cross-country variation. Country-level breakdowns (Figures 2–3) reveal that the pooled average hides substantial per-country heterogeneity. India leads in *computer programming* (7.6%), *mathematical calculation* (3.1%), and *tutoring/teaching* (9.3%), consistent with India’s large IT sector and the OpenAI Signals finding that Indian users are roughly three times above the global median for coding queries [28]. Nigeria dominates in *personal writing and communication* (12.7%) and shows the strongest engagement with text editing and critique. Brazil has the highest *health/fitness/beauty/self-care* share (7.7%, nearly double India’s) and the lowest programming share (2.4%). Pakistan leads in *translation* tasks and matches Nigeria in *personal writing or communication* (12.7%).

5.1.2 Comparison with OpenAI global averages. Table 3 compares the pooled coarse-topic shares across our four countries with those reported by Chatterji et al. [12]. Practical Guidance is the single largest domain (28.8%), closely matching the global share of 29.0%. The most informative comparisons are the deviations: our sample over-indexes on *Seeking Information* (+4.0pp) and *Writing* (+2.5pp) and under-indexes on *Technical Help* (−5.5pp) and *Multimedia* (−2.0pp). At the fine-grained level, the five most common categories are *specific information* (19.3%), *how-to advice* (11.3%), *tutoring or teaching* (8.9%), *personal writing or communication* (8.5%), and *edit or critique provided text* (6.3%).

Table 3. Coarse topic distribution: our four-country sample vs. OpenAI’s reported global averages [12].

Topic	OpenAI global (%)	Ours (%)	Diff (pp)
Practical Guidance	29.0	28.8	−0.2
Seeking Information	18.0	22.0	+4.0
Writing	19.0	21.5	+2.5
Technical Help	14.0	8.5	−5.5
Self-Expression	5.0	3.9	−1.1
Multimedia	7.0	5.0	−2.0
Other/Unknown	8.0	10.2	+2.2

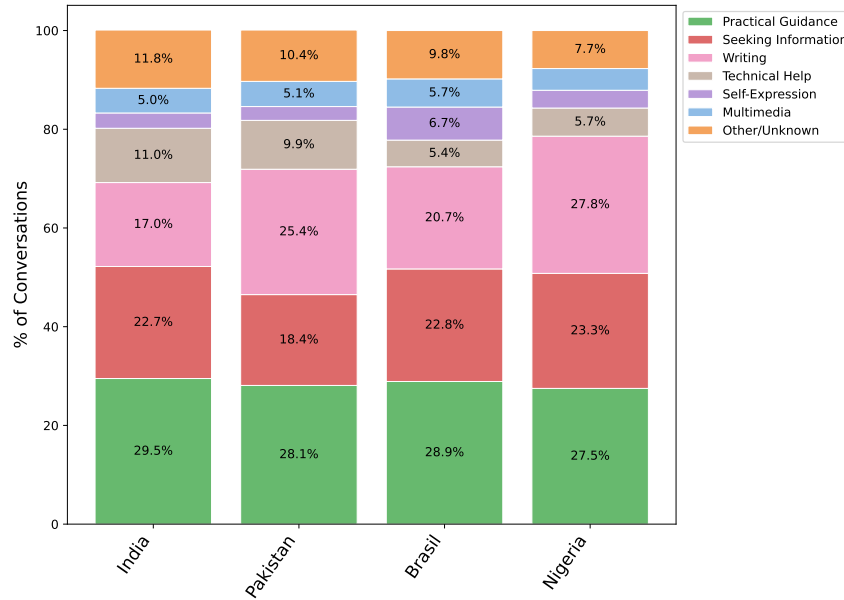


Fig. 2. Coarse topic distribution by country. Practical Guidance, Seeking Information, and Writing dominate across all four countries, collectively accounting for over 65% of conversations.

5.1.3 *Gender and age conditioning.* We emphasize up front that the gender and age composition of our sample reflects the convenience sample on Clickworker recruitment channel rather than the underlying ChatGPT user base; we therefore interpret gender- and age-*conditional* usage patterns, not gender or age composition, as findings.

Figure 4 shows that male users over-index on *Technical Help* (driven by computer programming) in all four countries, while female users over-index on *Practical Guidance* and *Self-Expression*, with the largest female skews in health/self-care, tutoring, and personal writing. The within-topic gender breakdown (Figure 30) is consistent with these divergence patterns. This conditional pattern is stable across all four countries and directly matches, at the individual-conversation level, the name-inferred pattern reported by OpenAI for India [28].

On age (Figure 31), the dominant cohort varies by country: 18–25 year olds contribute the largest share of conversations in India and Pakistan (with particularly pronounced representation in *Technical Help*), while the 26–35 cohort is largest in Nigeria and Brazil. The 36+ cohort is consistently the smallest in India and Pakistan but substantial in Brazil (where older users are well represented across all topics). In India, this youth dominance is consistent with OpenAI [28], which reports that 18–24 year olds account for nearly half of Indian ChatGPT messages.

5.1.4 *Temporal evolution.* Figure 5 shows the share of each coarse topic over time in our four countries. The early period of ChatGPT adoption (2023) was dominated by *Writing* and *Technical Help* queries, which decline sharply (most dramatically in Brazil, where *Writing* briefly exceeded 40% before collapsing) while *Practical Guidance* and *Seeking Information* grow steadily and become the dominant categories by 2025. This convergence to guidance and information-seeking matches the late-2025 snapshot reported in OpenAI Signals for India [28]. A stacked-area version of the same data, which additionally overlays monthly conversation volume, is deferred to Appendix D (Figure 38).



Fig. 3. Top 15 fine-grained topics by country. Colour intensity indicates the share of conversations devoted to each topic in each country.

5.2 Unsupervised Topic Discovery: Culturally Specific Themes

A predefined 24-category taxonomy cannot, by construction, surface topics that the taxonomy designers did not anticipate. We therefore complement the supervised classifier with an unsupervised pipeline (BERTopic with Gemini embeddings) described in Section 4. To avoid confusion with the OpenAI “fine-grained” 24-category level, we refer to the per-country outputs of this pipeline as *topic clusters* (approximately 50 per country) and to the cross-country rollup as *themes* (ten categories).

5.2.1 *Country-specific topic clusters.* Figure 6 shows the top 15 topic clusters per country. Three observations stand out.

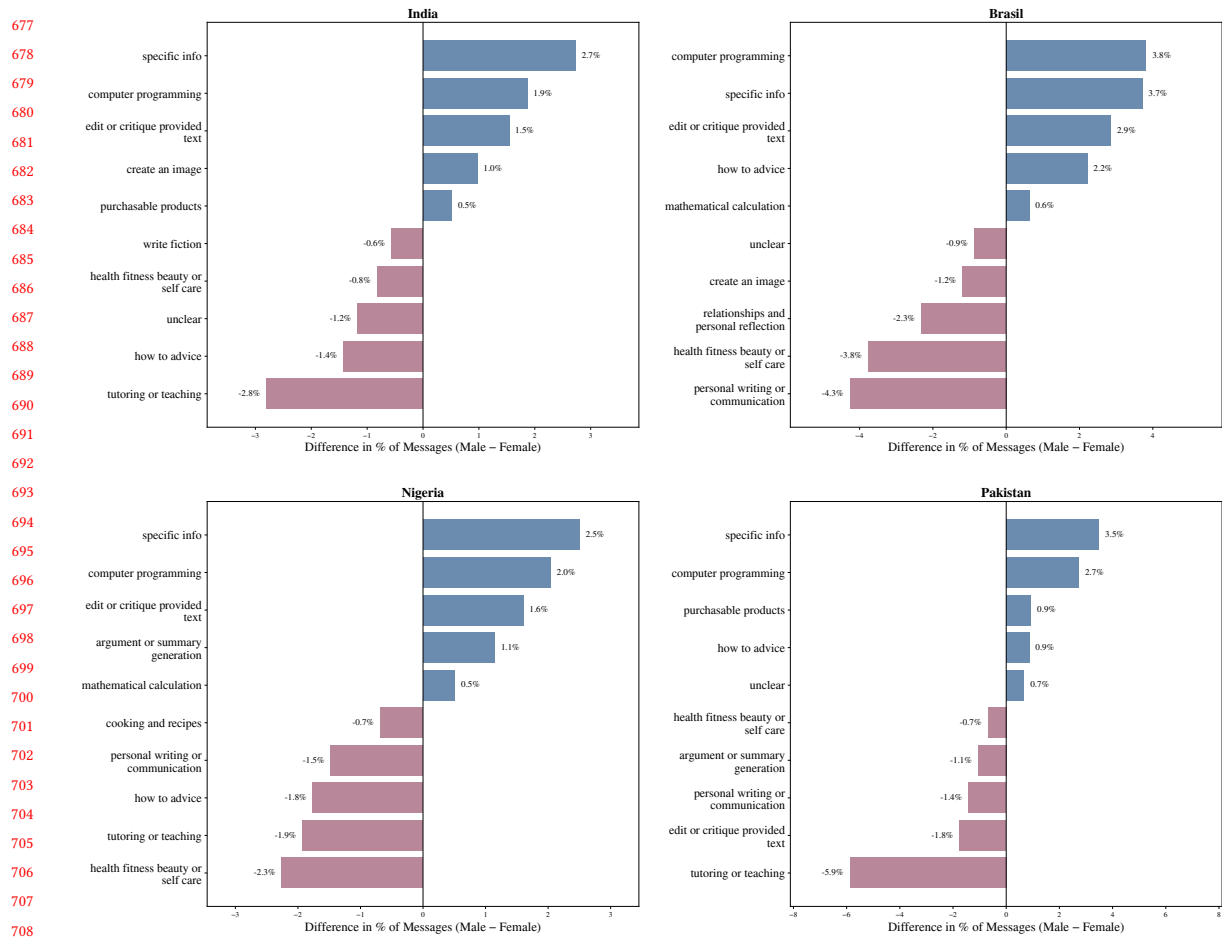


Fig. 4. Gender topic divergence: difference in topic share (Male - Female) per country. Blue indicates male over-representation; pink indicates female over-representation.

Health and wellness as a dominant theme. The OpenAI taxonomy subsumes these conversations under the smaller “health/fitness/beauty/self-care” fine category (7.7% in Brazil, 4.4% in India); the unsupervised clustering shows that once related wellness sub-topics—symptom interpretation, diet and fitness advice, mental-health queries—are aggregated semantically, health is the modal use case in two of our four countries. This growth in health-related queries is consistent with the trajectory documented by Karnam et al. [24], who find substantial increases in health and mental-health usage over time in their GDPR-donated ChatGPT corpus, and with ChatGPT serving as an accessible health-information substitute where formal consultation is expensive or uneven.

Culturally specific clusters that the OpenAI taxonomy does not have a name for. Religious clusters appear only in Nigeria (Religion, Literature & Philosophy, 5.3%) and Pakistan (Religious Queries, 2.8%), and are entirely absent from India’s and Brazil’s top-15 lists. Urdu-English translation is the single most prevalent cluster in Pakistan (7.2%); Portuguese-English translation is a top-5 cluster in Brazil (5.7%). Self-reflection and emotional conversations (5.7%) appears as a top-5 Brazilian

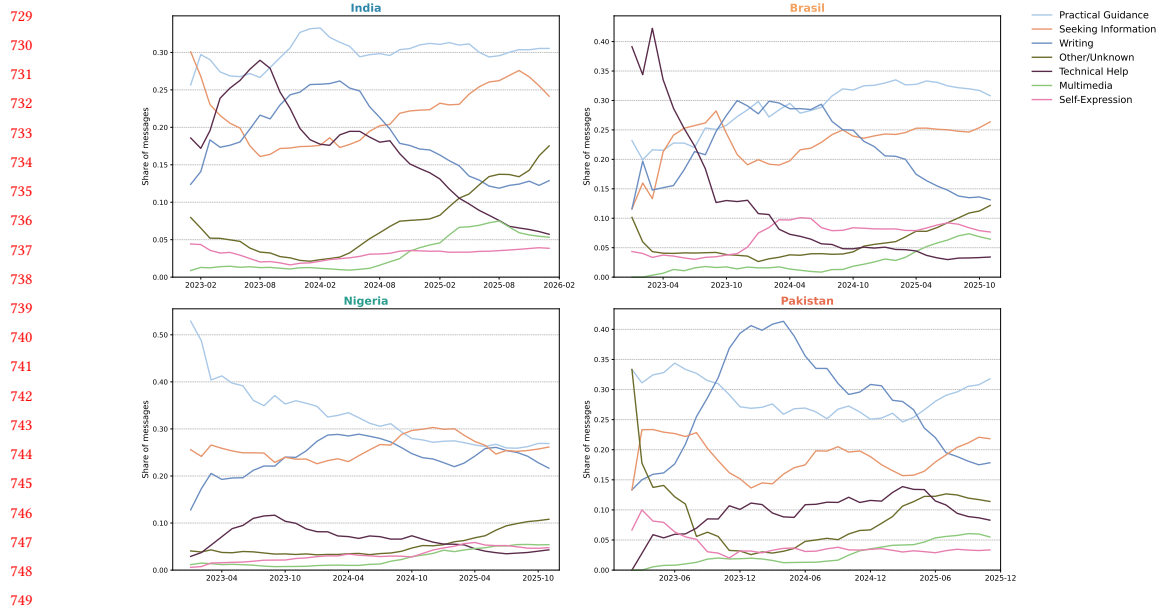


Fig. 5. Share of coarse topics over time, non-work conversations only.

cluster, unique among our four countries. *Online earning* and *YouTube monetization strategies* are top clusters in India (4.1% and 3.5%) and also surface prominently in Pakistan (5.0%). The prevalence of these clusters is likely amplified by the Clickworker recruitment channel (participants are, by definition, active gig-economy workers), but the fact that the same digital-entrepreneurship signature does *not* appear in Brazil or Nigeria (who were recruited through the same channel) suggests this reflects genuine cross-country differences in how ChatGPT is used as a resource for online income generation, rather than a pure sampling artifact.

Cross-country convergence on digital-entrepreneurship themes. Grouping the topic clusters into ten broad themes that span all four countries (Figure 7) shows that *Finance/Earning* is strikingly concentrated in India (17.7%) and Brazil (11.6%) but nearly absent in Nigeria (3.3%). *Programming/Tech* is most prominent in India (21.0%) and least in Brazil (8.3%). *Writing/Creative* is largest in Nigeria (11.6%) and Pakistan (8.6%). *Religion* is exclusively a Nigerian/Pakistani theme. *Translation/Language* is highest in Pakistan (11.2%).

5.2.2 Gender conditioning (themes). Figure 32 shows that gender-conditional patterns under the unsupervised lens are *sharper* than under the OpenAI taxonomy, because the clusters are thematically tighter. Male users over-index on *Programming/Tech* in every country and on *Finance/Earning* in India and Pakistan; *Religion* is a distinctly male theme in Nigeria and Pakistan. Female users over-index on *Health/Medical* in every country, and on *Content Creation* in Nigeria and Brazil. *Education/Academic* is the most female-skewed theme in Nigeria and Pakistan, suggesting that women in these countries disproportionately use ChatGPT as an academic support tool.

5.2.3 Age conditioning (themes). Figure 33 shows a clear generational gradient. The 18–25 cohort over-indexes on *Education/Academic* and *Programming/Tech* in India, Nigeria, and Pakistan, and on *Content Creation* and *Health/Medical* in Brazil, a student-oriented profile. The 26–35 cohort distributes most uniformly across themes. The 36+ cohort has

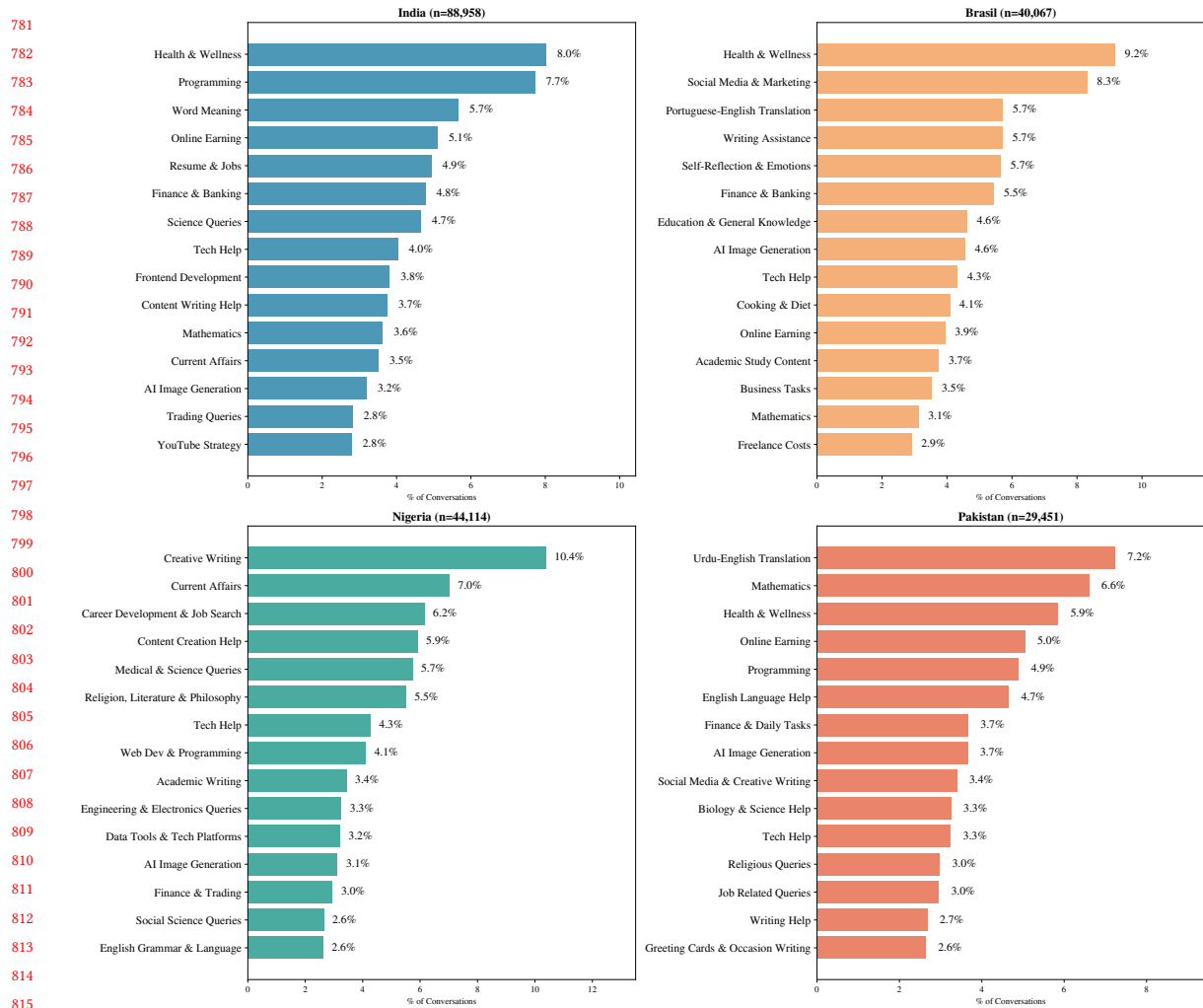


Fig. 6. Top 15 unsupervised topic clusters per country. The pipeline surfaces detailed information-seeking behaviour (e.g. the prevalence of Health and Wellness) and culturally specific clusters (e.g. Nigeria’s religious topics, Brazil’s self-reflection conversations, Pakistan’s Urdu–English translation needs, India’s digital-entrepreneurship themes) that the 24-category OpenAI taxonomy absorbs into generic buckets.

the most distinctive profile: in India it over-indexes strongly on *Finance/Earning* and *Content Creation*; in Nigeria and Pakistan on *Religion* and *Current Affairs*; in Brazil on *Job/Career* and *Writing/Creative*. Taken together, these patterns describe a gradient from education and skill-building among younger users toward financial, professional, and civic topics among older users, visible across otherwise very different national contexts.

5.2.4 Temporal evolution (themes). Figure 34 shows the monthly evolution of the ten themes. The broad trends mirror the OpenAI-taxonomy view: early dominance of programming and writing themes gives way to a more diversified distribution where health, education, and finance grow as shares of monthly volume.

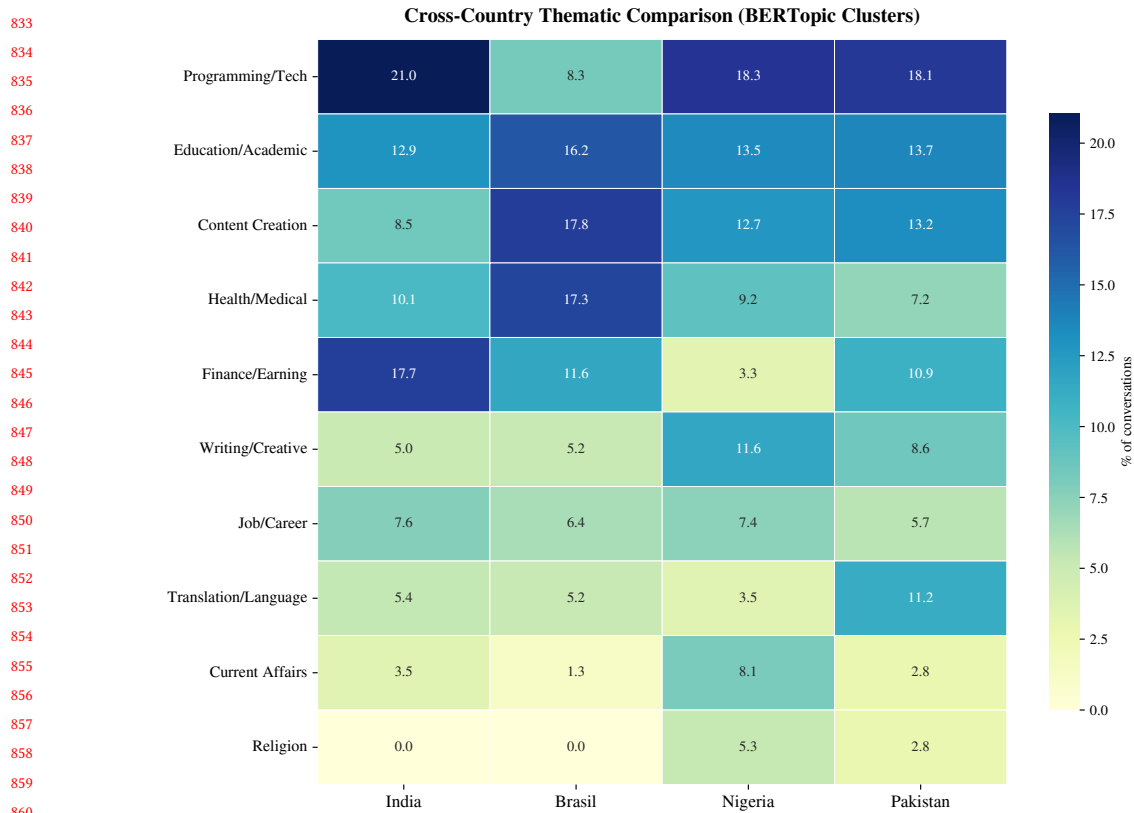


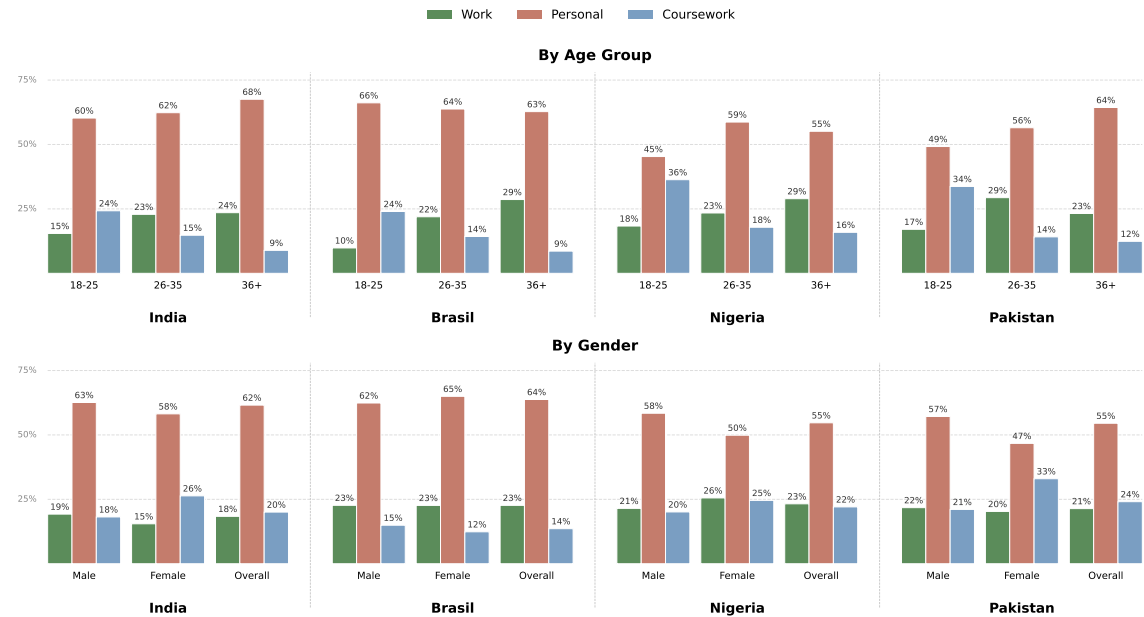
Fig. 7. Heatmap of ten broad themes (cross-country rollup of the unsupervised clusters) across countries. Colour intensity indicates the share of conversations per country.

5.3 Work, Coursework, and Personal Use

Following the task-purpose framework in Anthropic’s Economic Index [3], we label each conversation as work-related, coursework-related, or personal. The classifier was applied identically to all four countries. Personal use dominates everywhere: 61.5% of conversations in India, 63.7% in Brazil, 55.0% in Nigeria, and 55.4% in Pakistan. Work and coursework are of roughly comparable prevalence in the pooled sample (20.8% work vs. 19.6% coursework), though Brazil is an outlier with a substantially lower coursework share (13.6%) consistent with its older user base, and Pakistan has the highest coursework share (23.1%) consistent with its very young cohort. The 18.5% work share in our Indian subsample is noticeably lower than the 27% reported for global ChatGPT usage by Chatterji et al. [12], again consistent with the younger demographic composition of our sample and its larger share of students.

Figure 8 shows two conditional patterns worth flagging. First, as expected, the share of coursework drops monotonically with age in every country (from ~24–36% among 18–25 year olds to ~9–16% among 36+), while the share of work generally rises (from ~10–18% to ~23–29% in India, Nigeria, and Brazil; Pakistan is the exception, where work peaks in the 26–35 cohort before declining). Second, and more striking, female users use ChatGPT for coursework at substantially higher rates than male users in three of four countries: 26.1% vs. 18.2% in India, 32.8% vs. 19.7% in Pakistan, and 23.5% vs. 20.2% in Nigeria. Brazil is the exception (12.4% female vs. 14.9% male), in line with its distinct older, more

885 female-skewed, more work-oriented profile. The India/Pakistan/Nigeria pattern is consistent with ChatGPT serving as
 886 an academic support resource that women in these countries lean on disproportionately, complementing the female
 887 over-representation on the *Education/Academic* theme we reported under the unsupervised lens.
 888



912 Fig. 8. Distribution of task purpose (work, coursework, personal) by age group (top) and gender (bottom), by country.

913

914

915 **5.3.1 Topic distribution conditioned on task purpose.** Figure 9 quantifies the topic composition of work vs. non-work
 916 conversations under the OpenAI taxonomy. The first striking feature is that *Writing* and *Technical Help* combined
 917 account for roughly 45–50% of work-related conversations in every country (44.6% in India, 48.3% in Nigeria, 47.3% in
 918 Brazil, 50.0% in Pakistan); within that pair, *Writing* alone is consistently the single largest work bucket, accounting
 919 for 28–42% of work conversations. Even at the end of 2025, in other words, ChatGPT as a workplace tool is still
 920 predominantly a writing-and-editing tool for our four countries; the coding assistant framing that dominates popular
 921 discourse is secondary. The second feature is the mirror image on the non-work side: *Practical Guidance* is consistently
 922 the largest non-work bucket (28–31% across countries), followed by *Seeking Information* (20–26%), together covering
 923 more than half of non-work conversations. *Self-Expression* is nearly three times more prevalent in non-work than in
 924 work (3–8% vs. 1–3%), and only shows up meaningfully in the personal setting. The cross-country stability of this
 925 pattern (writing-heavy work, guidance-and-information-heavy personal use) mirrors the OpenAI Signals finding for
 926 India [28].
 927

928

929

930 Conditioning the unsupervised topic clusters on task purpose (Figures 10–11 and 35) makes the content of these
 931 work/personal conversations concrete. The writing-heavy work profile surfaced above resolves, under the unsupervised
 932 lens, into a small number of recurring work themes: *CV and cover-letter drafting*, *job-application and interview preparation*,
 933 *professional email and correspondence*, and *content creation for social and short-video platforms*. These are the specific tasks
 934 behind the “Writing at 28–42% of work” finding. *Online earning* appears as a top-5 work cluster in India, Pakistan, and
 935
 936 Manuscript submitted to ACM

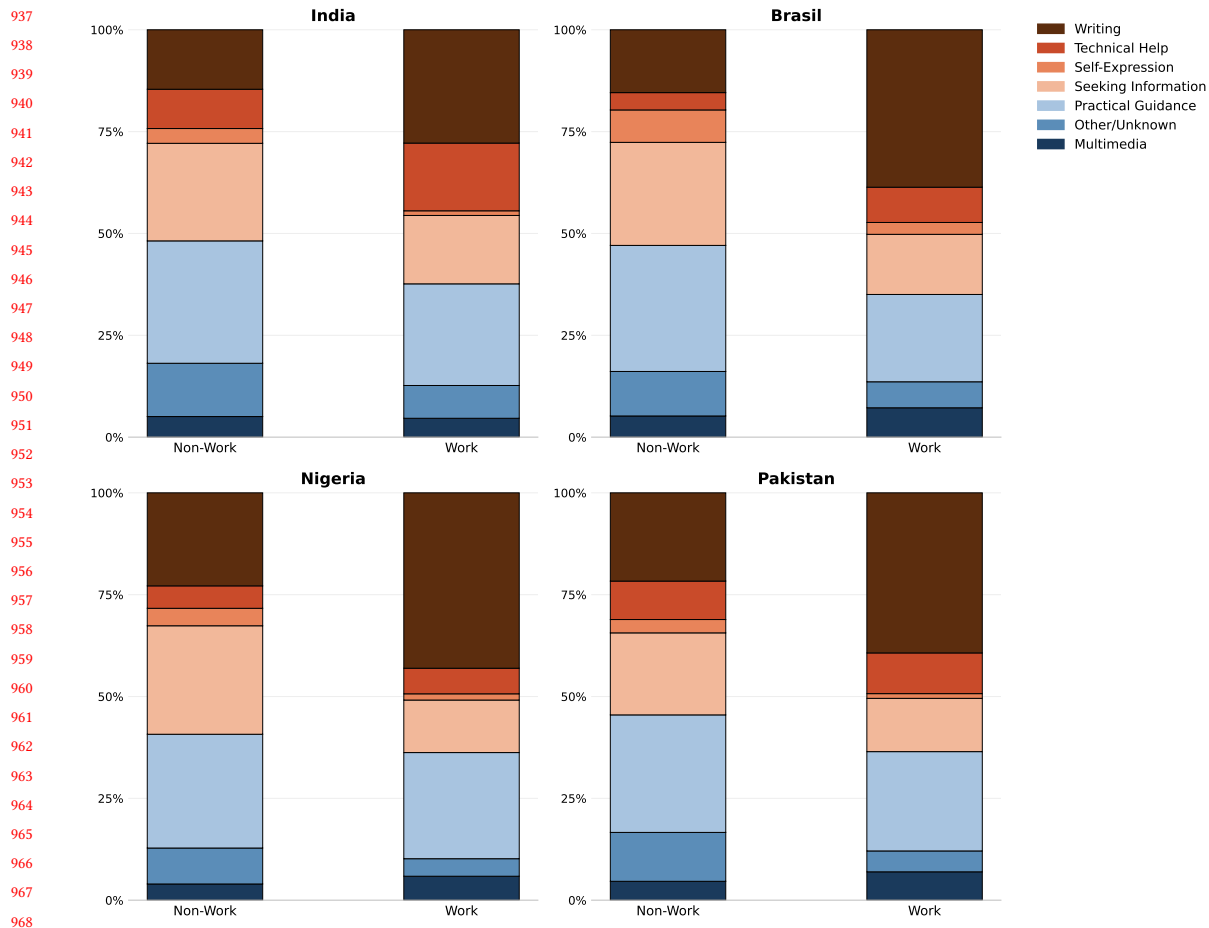


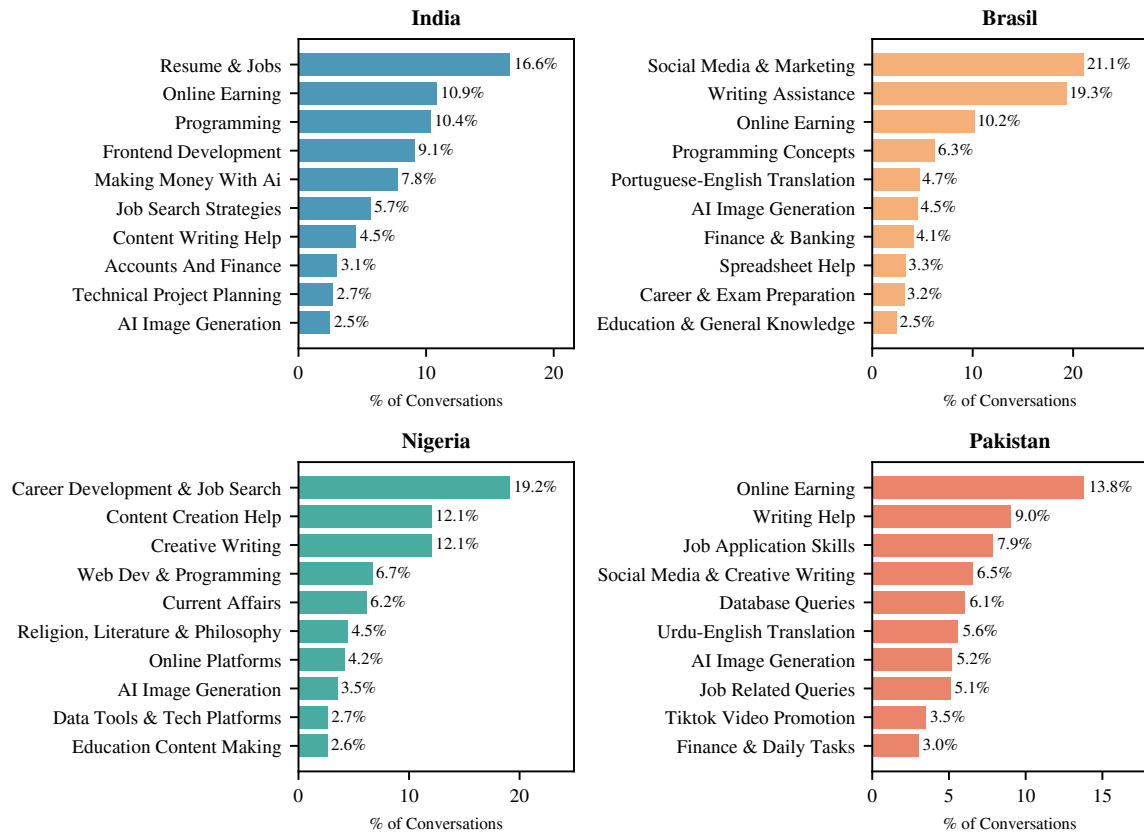
Fig. 9. Share of coarse topics for work-related vs. non-work conversations, by country.

Brazil, suggesting that across these economies users frequently turn to ChatGPT to navigate precarious or entrepreneurial labour-market conditions. Resume building and job applications dominate work-related use in India, while career development and content creation lead in Nigeria. Programming and technical tasks appear consistently but are never the largest work cluster in any single country, reinforcing that the technical-help slice of work is narrower and more targeted than the writing slice.

Personal conversations (Figure 11) are anchored by *Health and Wellness* in India and Brazil (where it is the top personal theme) and Pakistan (second). Beyond health, personal use is markedly more culturally differentiated: Nigeria stands out for creative writing and current affairs; Pakistan for Urdu–English translation and religious queries; Brazil for self-reflection and emotional support, suggesting a more expressive, interpersonal mode of engagement that is largely absent from the other three countries.

Coursework conversations (Figure 35) are dominated by STEM subjects in every country. Programming and mathematics lead in India and Pakistan; Nigeria concentrates on academic writing, engineering, and data tools; Brazil is

989 distinctive in its large share of general-knowledge queries and Portuguese–English translation, pointing to language
 990 support as a significant driver of educational use.
 991



1022 Fig. 10. Top 10 unsupervised topic clusters for work-related conversations, by country.
 1023
 1024

1025 5.3.2 *Temporal evolution of task purpose.* Figures 12–13 show that India, Nigeria, and Pakistan exhibit a remarkably
 1026 stable work/non-work split throughout the observed period, with non-work conversations consistently accounting for
 1027 75–85% of the total. Brazil is the notable exception, beginning in early 2023 with an almost even split between work
 1028 and non-work usage before converging to the ~80% non-work share seen elsewhere by mid-2024. The stability of the
 1029 Indian split is consistent with OpenAI [28], which reports that the shift toward non-work dominance in India occurred
 1030 before mid-2024 and has remained steady since.
 1031
 1032

1033 5.4 Asking, Doing, Expressing

1034 We apply the three-way *Asking/Doing/Expressing* intent classifier introduced by Chatterji et al. [12] (*Asking* = seeking
 1035 information or decision support; *Doing* = requesting the model to execute a task; *Expressing* = using the model for
 1036 reflection or emotional communication). Approximately 1.5–2.8% of conversations per country received malformed
 1037 classifier output and are excluded from ADE analyses; all percentages below are computed over valid labels only.
 1038
 1039

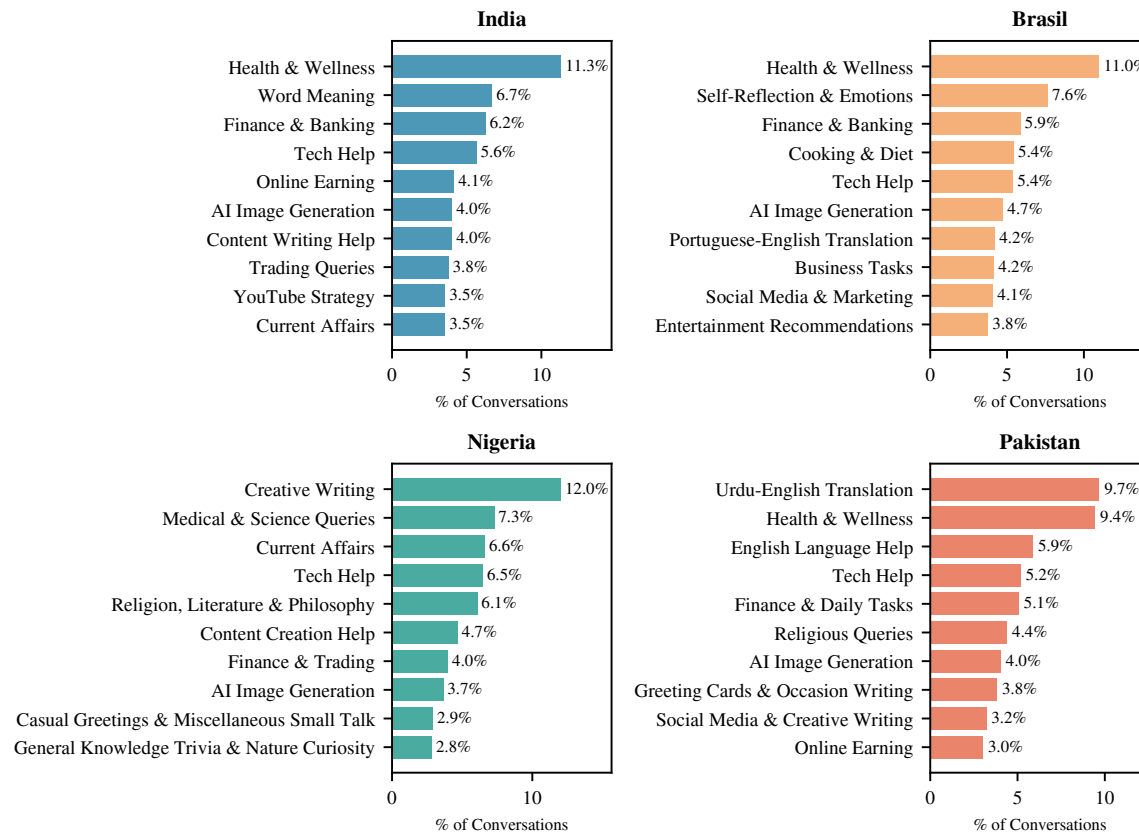


Fig. 11. Top 10 unsupervised topic clusters for personal conversations, by country.

Figure 36 shows the distribution conditioned on task purpose (work vs. non-work), and Figure 14 shows the monthly trend.

The two comparisons tell a consistent story. Work-related conversations are disproportionately *Doing* (drafting, transforming, or executing tasks) while non-work conversations are disproportionately *Asking* and *Expressing*. The shift is most pronounced in Brazil, Nigeria, and Pakistan, where *Doing* becomes the plurality or outright majority of work conversations (50.2%, 52.5%, and 49.1% respectively), while *Asking* drops by 13–18 percentage points from its non-work level. India shows a directionally consistent but smaller shift: *Asking* remains the plurality even in work (42.2%), with *Doing* at 38.7%. Across all four countries, the *Expressing* share declines from non-work to work, consistent with the OpenAI Signals finding for India that work conversations are more task-oriented [28].

Two points are worth emphasizing across these temporal trends. First, although *Asking* is the dominant mode in every country throughout the observation period, its share has gradually declined as *Doing* has grown, most visibly in India and Nigeria, where *Doing* rises from below 20% in early 2023 to 30–35% by late 2025. Chatterji et al. [12] report a similar global gradient, with *Doing* reaching roughly 40% of messages by 2025; our four-country corpus reproduces this trajectory, suggesting that as users mature on the platform they progressively delegate more autonomous task execution to the model. Second, there is meaningful cross-country heterogeneity even though the relative ordering (*Asking* >

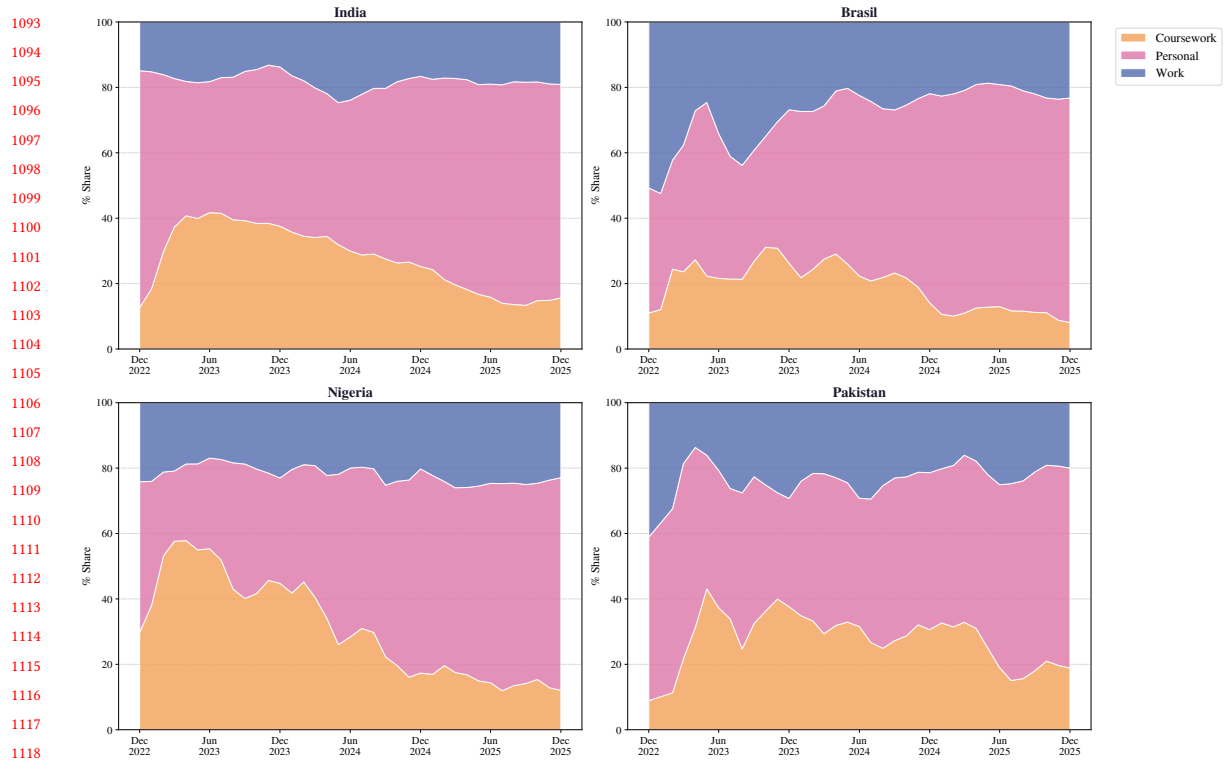


Fig. 12. Temporal share of work, coursework, and personal conversations (3-month rolling average).

Doing \gg *Expressing*) is consistent: India and Nigeria show the most pronounced shift toward *Doing*, Brazil maintains the highest share of *Expressing* throughout, and Pakistan’s *Doing* share remains smaller and more volatile. Perhaps most interestingly, the *Expressing* share (though small in absolute terms) exhibits a gentle upward drift over time across all four countries. This is consistent with two independent lines of evidence: Karnam et al. [24] document rising social framing and companion-like treatment of ChatGPT over time, and Fang et al. [19] find that heavier users engage in proportionally more reflective (as opposed to purely transactional) exchanges. We flag this as an underexplored trend worth following in future work (Section 6): if users are using ChatGPT for reflective, affective, or companion-like conversation at a slowly growing rate, the downstream implications for well-being, emotional-support infrastructure, and product design are substantial.

5.4.1 What *Expressing* actually contains: a qualitative reading. The *Expressing* category is the most interesting of the three, and two pieces of evidence sharpen what it actually contains. The first is linguistic, and we report it in full in Section 5.5: *Expressing* conversations are conducted in the country’s dominant language at substantially lower rates than *Doing* conversations in every country (e.g. 88.6% vs. 68.7% in India; 86.7% vs. 73.9% in Pakistan), with *Asking* falling between the two. Task delegation often involves code, structured instructions, or professional templates that draw on the dominant language even in non-Anglophone contexts; reflective and emotional exchanges, by contrast, are precisely the contexts in which users revert to the language they think in.

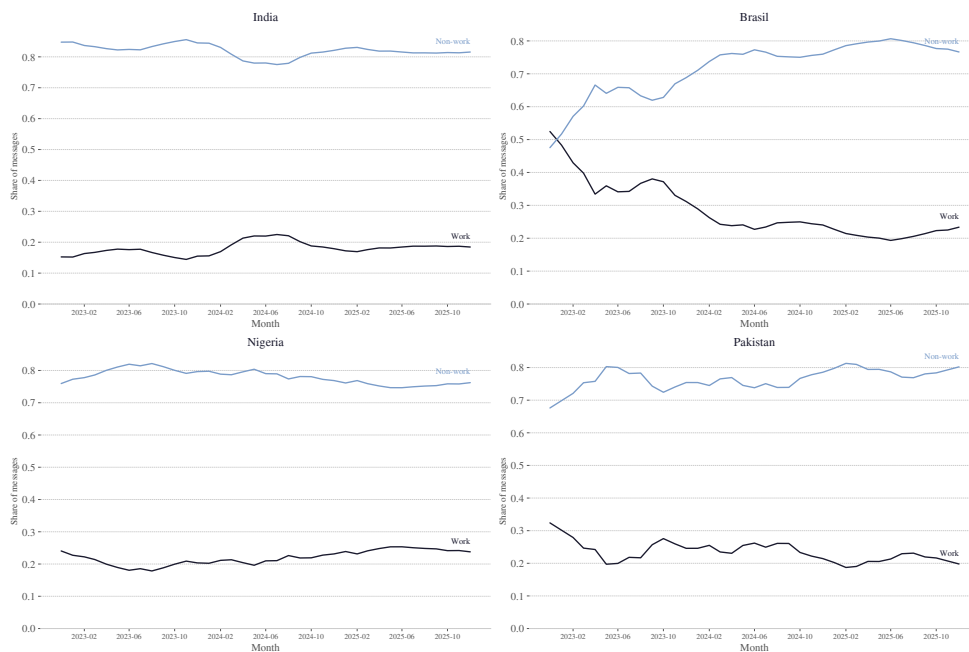


Fig. 13. Share of work-related vs. non-work conversations over time, by country.

The second is qualitative. To see what *Expressing* conversations actually contain, one of the authors hand-coded around 100 *Expressing* conversations per country using the message text plus the surrounding conversation context. Two things stood out. First, very few of these are pure emotional venting; almost every one contains a concrete request. Second, the request and the disclosure cannot be cleanly pulled apart. The topic taxonomy would file most of these as *Practical Guidance* or *Seeking Information*, which is correct at one level but loses the texture of what the user is actually asking. A 23-year-old in India narrates their long-distance relationship at length, says his 19-year-old girlfriend sends him skincare and a red horn wallet and “feels like a world” to him though they have never met, and then asks whether the age gap is a good match in India. A Pakistani user lists his symptoms (“confused mind, stiffness in brain, concentration problem”) before asking what clozapine, the drug his doctor has just prescribed, is for. Another Pakistani user, deep in an exam cycle, asks whether the fact that his brain only “works” near deadlines is “practically good or bad.” A Brazilian user pastes 1 Thessalonians 2:11–13 and then discloses that she is in a training program to become a pastor’s wife and has just realized her autism is making the role impossible; her question is no longer about the verse. A Nigerian user writes that she paid forty-three thousand naira to an online “foodstuff seller” who has gone silent and asks what to do. In all of these, the disclosure is what makes the answer specific. A search engine could field the surface question—what is clozapine, what does 1 Thessalonians 2:11 mean, what should I do about a missing order—it could not field the question with the disclosure attached, which is the question users actually want answered. A subset of these conversations goes further still: users name the assistant (“Noor,” “bhai”), refer back to earlier sessions, type only “Upset” and accept the offer to “sit quietly,” or scold the model with profanity for botched accountancy. For these users the relationship has shifted past instrumental use, into something closer to a confidant.

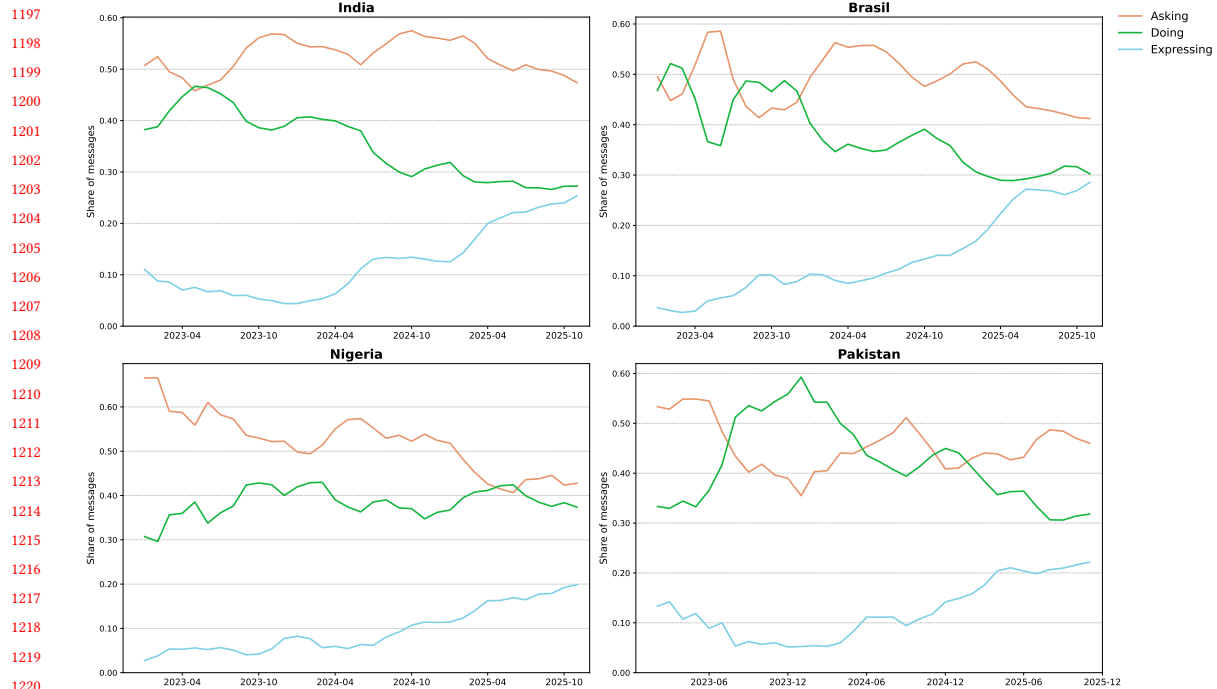


Fig. 14. Temporal shift in *Asking/Doing/Expressing* share (3-month rolling average) by country.

This qualitative pattern is consistent with the classifier-validation results we report in Appendix B: the *Asking/Doing/Expressing* classifier reaches substantial overall agreement with human labels ($\kappa = 0.55$), but the *Expressing* class has high recall (0.78) and low precision (0.35) against the human gold labels. The dominant error mode is the classifier flagging conversations as *Expressing* when an annotator would have called them *Asking* or *Doing*, precisely because the conversations contain enough self-disclosure to look expressive even when the user’s request is instrumental. The classifier’s noise is therefore consistent with the substantive finding: pure *Expressing* is rare; the dominant mode is hybrid.

5.5 Raw-Metadata Descriptives: Model Adoption, Tools, and Engagement

Unlike platform reports, our conversation exports expose raw client-side metadata: the model version used for each turn, tool invocations (web search, code interpreter, image generation), prompt and response lengths, conversation depth (turn counts), and Free vs. Plus subscription status. We summarize these below.

Model version adoption. Figure 15 shows three distinct eras. The initial Text-Davinci model family (GPT-3.5) was universal through early 2024. GPT-4o then displaced it within roughly two months of its mid-2024 launch, dominating at over 95% of conversations through mid-2025. The GPT-5 family (GPT-5, GPT-5.1, GPT-5.2) began displacing GPT-4o from mid-2025 onward, with GPT-5.2 accounting for the majority of conversations by early 2026. The rapidity of each transition largely reflects that 89% of our sample are Free-tier users who receive whatever model the platform sets as the default; model selection is structurally unavailable to them rather than actively declined.

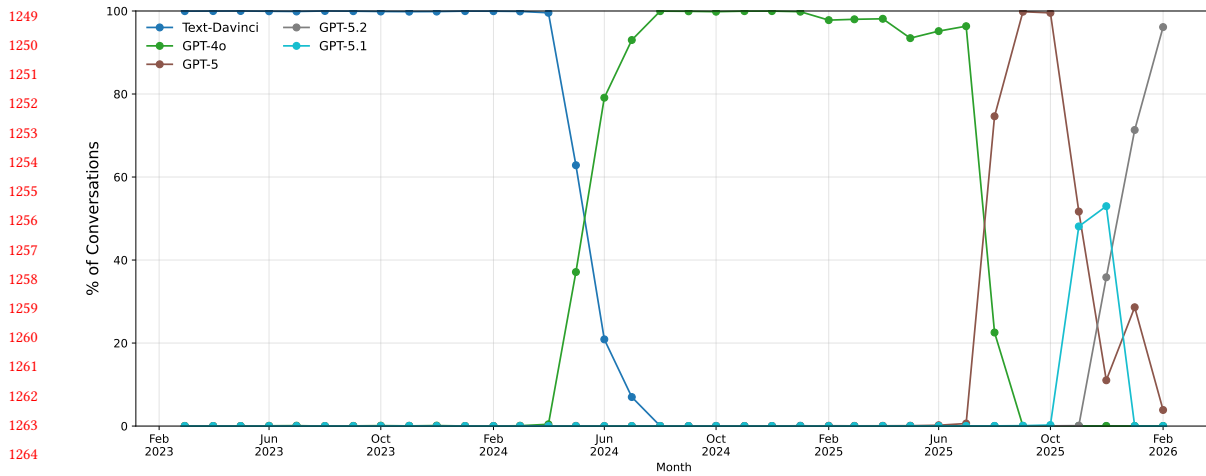


Fig. 15. Model version adoption over time. Three successive waves are visible: Text-Davinci (GPT-3.5), GPT-4o, and the GPT-5 family.

Tool and feature usage. Figure 16 shows feature adoption rates by country, recomputed from the raw JSON conversation data (authoritative user list only, $n = 1,252$). Web search is the most-used tool, appearing in 12.3–15.9% of conversations, with India (15.9%) and Brazil (15.6%) leading. Code interpreter usage is relatively uniform (3.4–4.1%) and image generation usage remains below 1% everywhere. These rates are modest overall, pointing to either limited feature awareness, limited relevance to user tasks, or free-tier access restrictions.

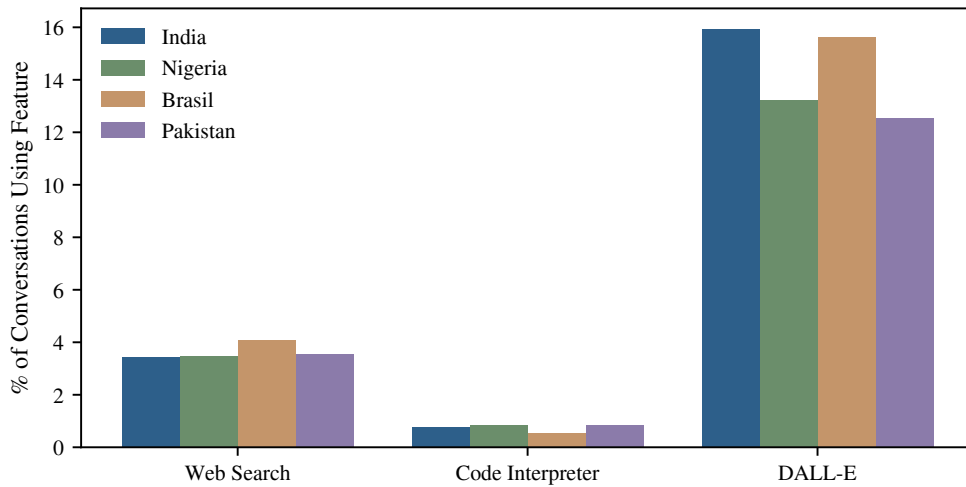


Fig. 16. Advanced feature adoption rates by country, computed over the raw conversation JSON for the authoritative 1,252-user cohort.

Free vs. Plus. Only 10.7% of users in our cohort (134 of 1,252) are ChatGPT Plus subscribers, and the distribution is strongly skewed toward India (116 of 134); Nigeria has only 1 Plus user and Pakistan only 3 (Table 4). Plus users

run deeper conversations, write longer prompts, and receive longer responses (Figure 17; all comparisons shown with bootstrap 95% CIs). The causal direction is ambiguous: users who invest in a paid tier may naturally use it more intensively, or the tier itself unlocks deeper interaction.

Table 4. Free vs. Plus user counts by country (authoritative 1,252-user cohort).

Country	Free	Plus
India	441	116
Nigeria	242	1
Brazil	232	14
Pakistan	203	3
Total	1,118	134

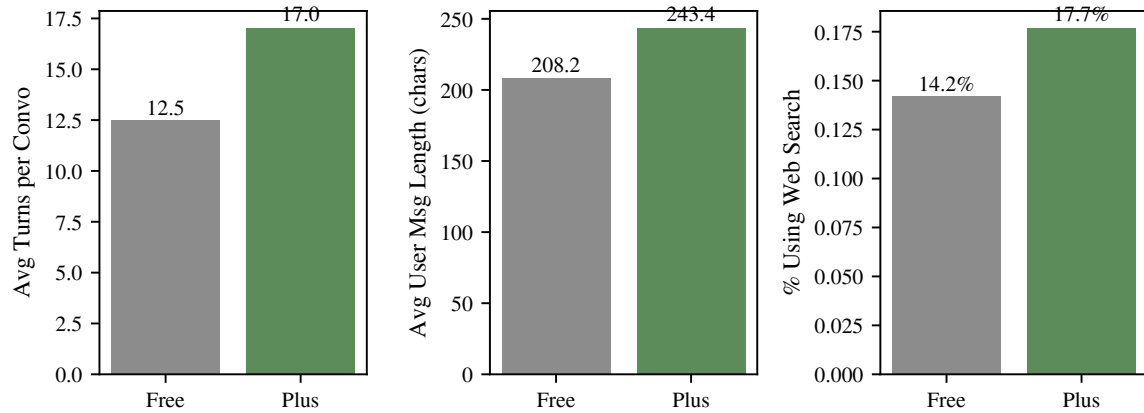


Fig. 17. Behavioural comparison between Free and Plus users (bootstrap 95% CIs).

Conversation depth and prompt/response lengths. Figure 18a shows the mean number of turns per conversation by country with bootstrap 95% confidence intervals. Mean conversation depth ranges from 11.6 turns in Nigeria to 14.6 in Pakistan; all CIs are tight, so the differences are statistically reliable despite being moderate in magnitude. The distributions are heavily right-skewed, so the mean reflects both the typical short exchange and the long-tail power users. Figure 18b shows mean prompt and response lengths per conversation with bootstrap 95% CIs. Pakistani users write the longest prompts on average, consistent with their higher share of technical and educational queries; response lengths are more uniform and consistently several times longer than prompts.

Adoption curves and demographic share of activity. Figure 37 shows monthly active users and conversation volume by country; India has the steepest trajectory. Figures 19–20 show gender and age shares of weekly active users, estimated via bootstrapping with equal samples per demographic group. All four countries show a slight increase in the female share of conversations over time, and in India the growth accelerates from mid-2025, matching the trend reported by OpenAI [28]. While the equal-sample bootstrap controls for headcount differences between demographic groups, cross-country level differences in activity share may still reflect the composition and engagement patterns of the

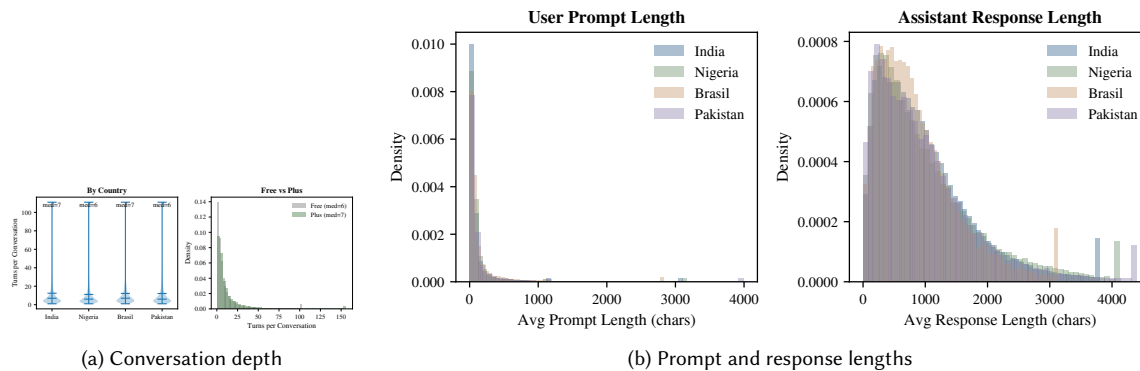


Fig. 18. (a) Mean turns per conversation by country; (b) mean prompt length (left) and response length (right) per conversation by country. All panels show bootstrap 95% CIs.

Clickworker sample rather than the broader ChatGPT user base; we therefore focus on *within-country trends over time*, which are consistent with platform-level reports.

Figure 21 shows usage intensity by hour of day and day of week (all timestamps converted to each country’s local timezone). Indian and Pakistani users are most active in the evening and late night; Nigerian users peak during business hours; Brazilian users are comparatively uniform throughout the day. Weekend activity is slightly lower than weekday activity across all countries. We report these as descriptive context rather than as substantive findings.

Conversation language. We run language detection (langdetect, deterministic seed) on the concatenated user messages of each conversation, using the first 5,000 characters and requiring a minimum of 20 characters for detection. Overall, English is the dominant language in India (80.0%), Nigeria (92.9%), and Pakistan (83.1%); Portuguese dominates in Brazil (83.2%). Hindi accounts for 1.8% of Indian conversations and Urdu for 2.3% of Pakistani ones; the remainder is a long tail of low-frequency languages.

The dominant-language share varies systematically with both task purpose and conversational intent (Figures 22–23). *Work* and *coursework* conversations are conducted in the dominant language at substantially higher rates than *personal* conversations: in India, 90.4% and 90.1% vs. 73.3% English; in Pakistan, 91.3% and 91.3% vs. 76.6%; in Brazil, 87.6% and 86.6% vs. 81.0% Portuguese. Nigeria shows the same ordering but a smaller gap (97.0%/95.1%/90.4%), consistent with English being the primary language of formal communication in Nigeria. The pattern suggests that users switch to their native language primarily when using ChatGPT for personal needs, where the communicative stakes favour comfort over the terminological conventions of English-language technical and academic discourse.

The Asking/Doing/Expressing intent decomposition reveals a parallel gradient. *Doing* conversations (task delegation) have the highest dominant-language share in every country (88.6% in India, 96.2% in Nigeria, 85.2% in Brazil, 86.7% in Pakistan), while *Expressing* conversations (emotional, reflective, opinion-sharing) have the lowest (68.7%, 85.4%, 76.3%, 73.9%). *Asking* falls between the two. The gap between Doing and Expressing is 10–20 percentage points in India and Pakistan, and 9–11pp in Nigeria and Brazil. This is intuitive: task delegation often involves code, structured instructions, or professional templates that are English-dominant even in non-Anglophone countries, whereas reflective and emotional exchanges are precisely the contexts in which users revert to the language they think in. For Brazil, the drop in *Expressing* Portuguese share to 76.3% means that roughly one in four *Expressing* conversations is in neither Portuguese nor English, pointing to a linguistically diverse user base.

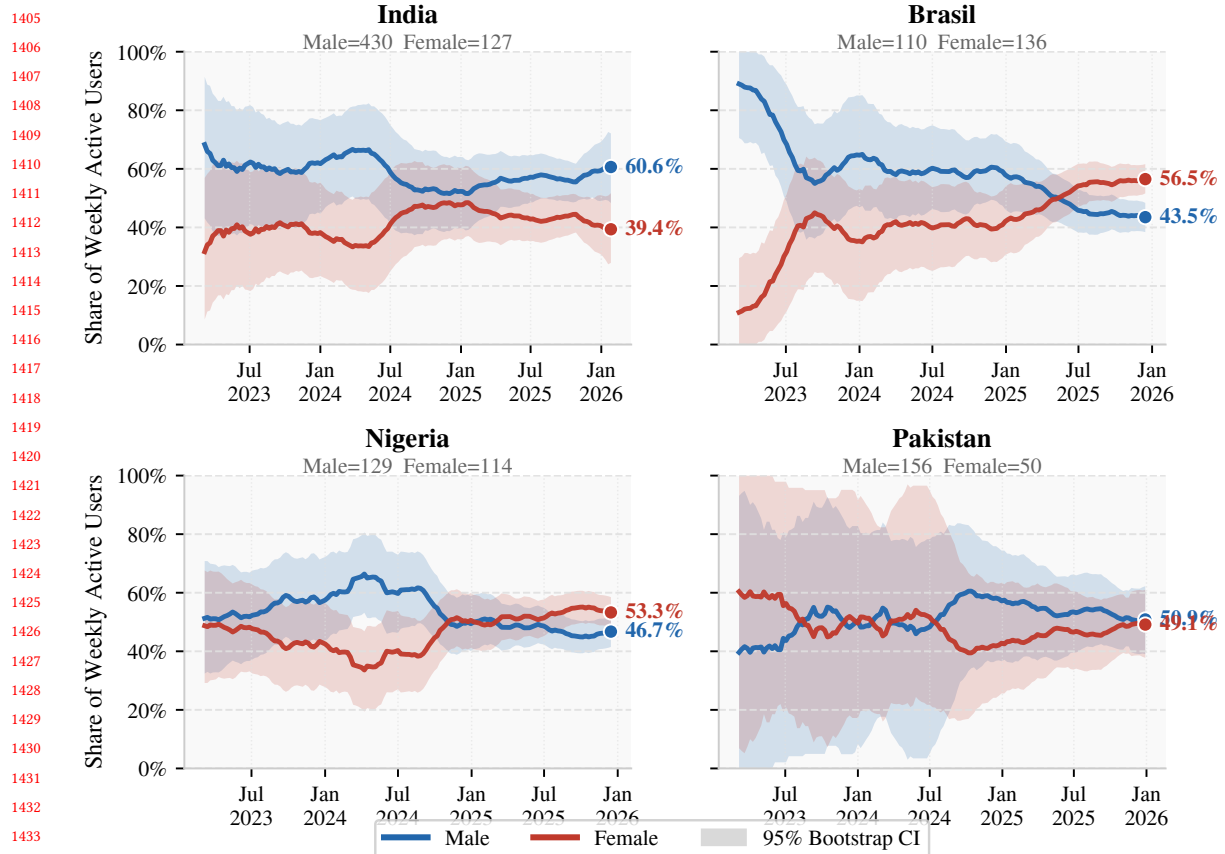


Fig. 19. Weekly-active-user share by gender and country (bootstrap CIs, 500 iterations).

Temporally, English share has declined in every country except Nigeria over the observation period (Figure 24). In India the decline is steepest: from 91.7% in 2023 to 77.7% in early 2026. Brazil shows the inverse trend for Portuguese (rising from 70.1% to 84.9%), reflecting a shift away from English and toward the local language as users become more comfortable conversing with ChatGPT in Portuguese. These trajectories suggest that ChatGPT's early user base in these countries skewed toward English-proficient individuals, and that as adoption broadened to include less English-dominant users, the linguistic composition of the conversation corpus shifted accordingly.

6 Discussion

Our results paint a picture of ChatGPT usage in the Global South that diverges in important ways from the productivity-focused narrative that dominates public and policy discourse around large language models. We organize this discussion around the broader implications of our findings, what they suggest for future research, and what they cannot tell us.

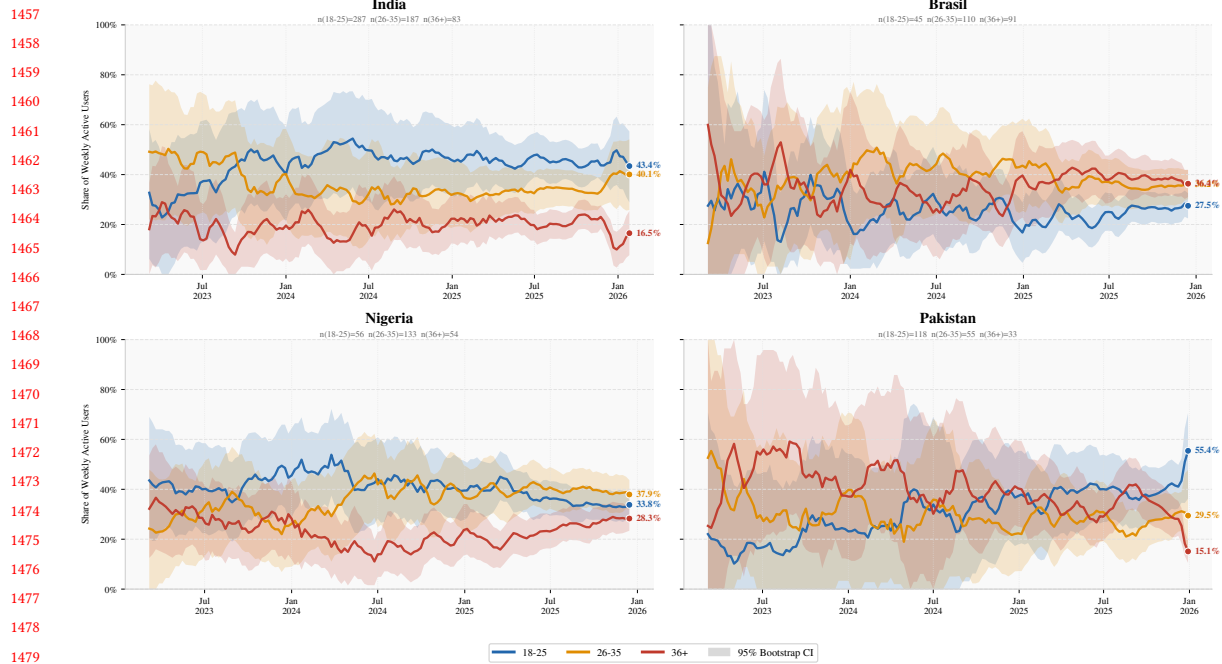


Fig. 20. Weekly-active-user share by age group and country (bootstrap CIs, 500 iterations).

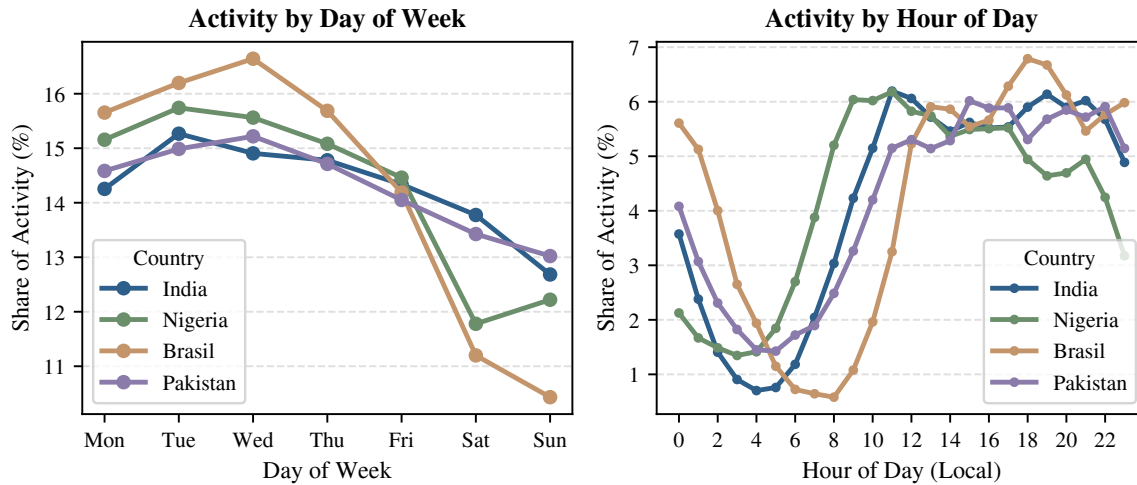


Fig. 21. Usage intensity by hour of day (left) and day of week (right) in local time, one line per country.

6.1 ChatGPT as Invisible Infrastructure

The single most striking pattern across our results (visible under every analytical lens we applied) is that ChatGPT in these four countries is overwhelmingly a *personal* tool. Personal conversations account for 55–64% of usage in every country; *Practical Guidance* and *Seeking Information* are the dominant coarse categories; and the unsupervised

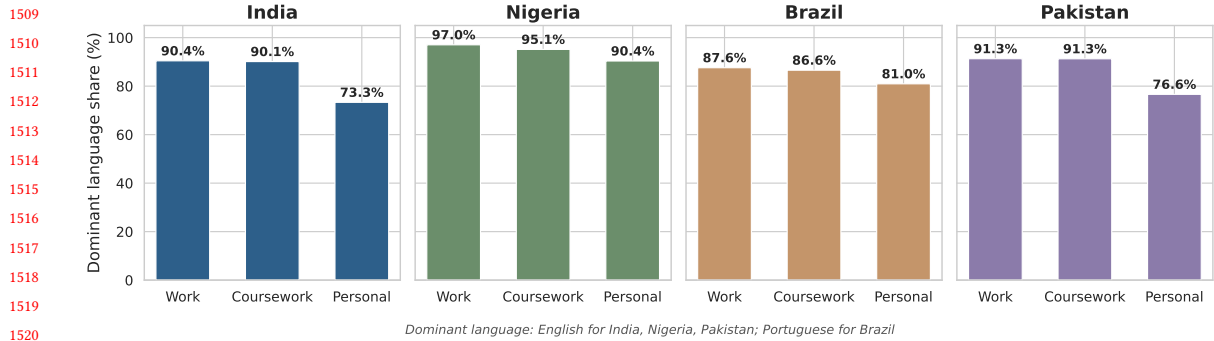


Fig. 22. Dominant-language share by task purpose (work, coursework, personal) and country. Dominant language is English for India, Nigeria, and Pakistan; Portuguese for Brazil. Personal conversations are substantially more likely to be conducted in a non-dominant language.

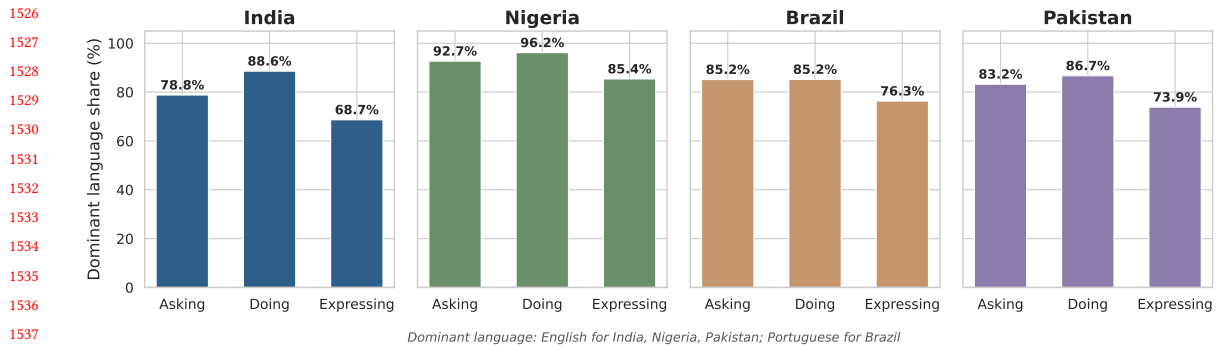


Fig. 23. Dominant-language share by conversational intent (Asking, Doing, Expressing) and country. Expressing conversations have the lowest dominant-language share in every country, consistent with users switching to their native language for reflective and emotional exchanges.

pipeline surfaces health and wellness, translation, and everyday problem-solving as the largest topic clusters. Even within work-related conversations, the dominant tasks are not the high-skill professional applications that receive the most attention (software engineering, data analysis, legal reasoning) but rather job-search support, CV and cover-letter drafting, and online-earning strategies, tasks at the boundary of formal employment and personal economic survival.

This matters for how we think about the economic value of conversational AI. The mainstream framing, reflected in Eloundou et al. [18] and Anthropic [3], centres on workplace productivity: how much faster can a programmer, writer, or analyst work with an LLM? That framing is important but structurally incomplete. A user who asks ChatGPT to interpret medical symptoms before deciding whether to pay for a doctor visit, or to diagnose why a household appliance is malfunctioning, or to draft a message to a landlord in a second language, is extracting real economic value, but it is value that does not appear in any standard measure of labour productivity. Economists recognize this as household production [8]: the non-market work that sustains daily life but is invisible to GDP accounting. Our data suggest that the primary mode of ChatGPT usage in the Global South falls squarely in this category. If so, productivity-oriented evaluations of AI impact will systematically undercount the technology's value in precisely the markets where it may

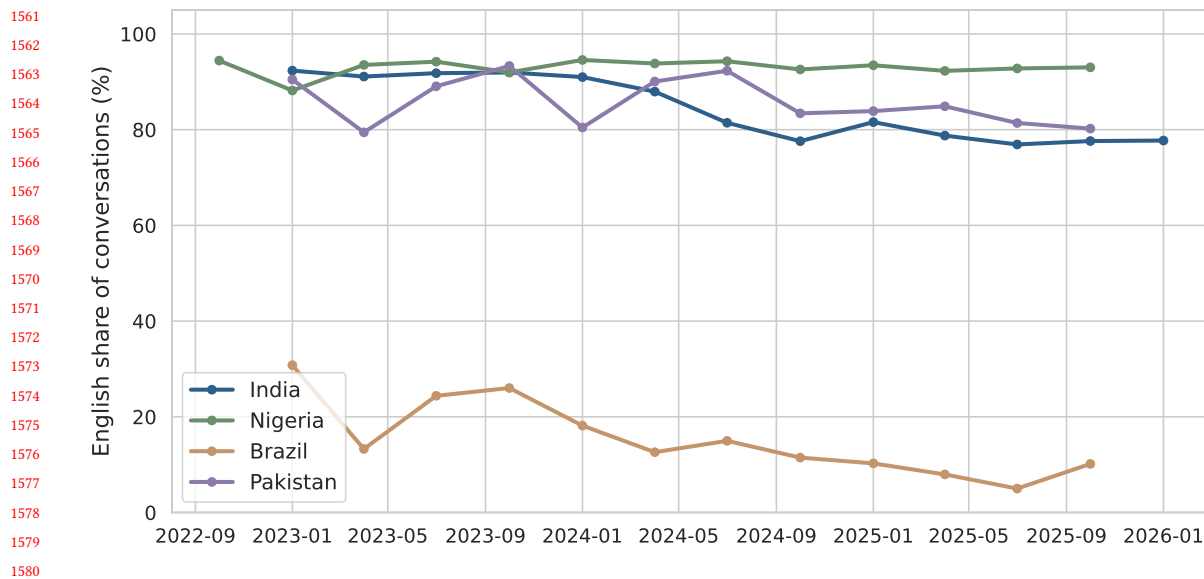


Fig. 24. English share of conversations by quarter and country. India shows a steady decline from over 90% in 2023 to under 80% by 2026; Brazil’s English share (not shown; Portuguese share rises from 70% to 85%) mirrors this trend in reverse.

matter most: markets where formal consultation (medical, legal, financial) is expensive relative to income, and where an always-available, zero-marginal-cost advisor fills a genuine gap.

The *content* of personal use reinforces the point: it is not frivolous. Health and wellness is a highly prevalent unsupervised theme overall in India (10.1%) and Brazil (17.3%), and a top-5 theme in Pakistan, in countries where the ratio of physicians to population is a fraction of that in the OECD. Translation is the top cluster in Pakistan and a prominent cluster in Brazil, bridging the gap between a user’s dominant language and the language in which information exists. The language-detection analysis (Section 5.5) reinforces this: personal conversations are 10–17 percentage points less likely to be in English than work or coursework conversations in India and Pakistan, and users are most likely to switch to their native language in *Expressing* (reflective, emotional) conversations, where the communicative stakes favour comfort over technical convention. A student in Pakistan who uses ChatGPT to understand a difficult concept in English (because no widely available Urdu explanation exists) is receiving a service with direct educational value. A gig worker in India who uses it to draft a competitive freelancing proposal is investing in income generation. These returns are real but invisible to the question “what percentage of knowledge workers use AI at their desks?” Any aggregate cost-benefit analysis that focuses exclusively on the workplace will miss most of the picture in these markets.

We cannot assess the quality of ChatGPT’s health advice from our data, and we make no claims that it is a substitute for professional care. But the sheer volume of health-related usage suggests that millions of users are already treating it as a first-line health-information resource. Understanding whether this substitution improves or harms health outcomes is an urgent empirical question that our data can motivate but not answer.

6.2 What Fixed Taxonomies Cannot See

A methodological lesson runs through our results: a taxonomy designed in one context will structurally fail to see the things that matter in another. OpenAI’s 24-category classification is well-designed for its purpose (stable, reproducible,

1613 directly comparable across billions of conversations) but it has no category for Islamic jurisprudence queries, Urdu–
1614 English code-switching, YouTube monetization strategy, or emotional self-reflection in Portuguese. These are not edge
1615 cases in our data; they are prominent clusters in their respective countries. The unsupervised pipeline finds them
1616 because it is free to discover whatever structure the data contains, rather than projecting onto a fixed vocabulary.
1617

1618 This observation generalizes beyond our specific context. Any study that classifies AI usage through a fixed taxonomy
1619 will inherit the blind spots of whoever designed that taxonomy. In a field where the most influential usage statistics come
1620 from platform operators in a small number of countries, this creates a systematic risk: the categories through which we
1621 understand AI usage worldwide are calibrated to the usage patterns of early adopters in high-income, English-speaking
1622 markets. Conversation-level access to the underlying data (the kind of access our study provides) is the only corrective.
1623
1624
1625

1626 **6.3 Conversational AI as Disclosure Plus Delegation**

1627 The intent results show a second kind of shift. Over time, users delegate more tasks to ChatGPT: *Doing* rises from
1628 below 20% in 2023 to 30–35% by late 2025, most visibly in India and Nigeria, matching the global trend reported by
1629 Chatterji et al. [12] and consistent with a maturing user base that has learned what the model can reliably do. Users
1630 begin by treating ChatGPT like a search engine and gradually learn to make it act.
1631

1632 *Expressing* is the more distinctive signal. By the end of our observation window it accounts for roughly a fifth of
1633 conversations in every country and is growing steadily, a mode of use prior population-scale information tools could
1634 not host. Adopting it requires a different behavioural shift than *Doing*: users have to learn they can disclose personal
1635 context, uncertainty, or vulnerability to a machine and expect a useful response, a trajectory aligned with Karnam et al.
1636 [24] and Fang et al. [19]. The qualitative reading (Section 5.4.1) shows the recurring pattern: users disclose personal
1637 context and then ask the model to do something with it—interpret a relationship, explain a medical worry, advise on a
1638 life decision, interpret a scriptural passage. This hybrid form, *disclosure plus delegation*, is what makes conversational
1639 AI different from search. A search engine can answer “what is clozapine”; it cannot easily answer the question after the
1640 user has attached symptoms, fear, or religious identity. Search trained users to minimize context; chatbots train them to
1641 do the opposite, because more context produces more useful answers. A subset of these conversations goes further
1642 still: users name the assistant, refer back to earlier sessions, or treat it as a confidant. Phang et al. [31] document this
1643 affective use at platform scale and show measurable well-being implications, and the earlier human–chatbot literature
1644 on Replika [34] anticipates the parasocial concerns raised by Turkle [37].
1645
1646
1647
1648

1649 The implications are double-edged. More context produces more relevant help, especially where users lack other
1650 resources. But the same context often involves health, mental health, finances, religion, relationships, or identity, and
1651 users are disclosing increasingly intimate material as they get more proficient with the tool. *Safety systems that treat*
1652 *emotional support and task completion as separate domains will miss the overlap, because the risky cases are often both*
1653 *at once*. The concern is sharpened commercially: search engines trained users to be terse, which incidentally limited
1654 what advertisers could learn from a query; chatbots train users to be expansive about exactly the material advertisers
1655 and data brokers most want, just as the platforms hosting these conversations begin to monetize them. The privacy
1656 floor is dropping as the commercial incentive to harvest disclosure rises. The language pattern reinforces the point:
1657 *Expressing* conversations are 10–20 percentage points less likely to be in English than *Doing* conversations in India and
1658 Pakistan (Section 5.5), making multilingual and culturally aware safety evaluation central rather than peripheral. And if
1659 users are turning to ChatGPT for companionship or emotional processing at scale, questions of parasocial attachment,
1660 displacement of human social ties, and the quality of AI-mediated emotional “support” become urgent, and they are
1661
1662
1663
1664

1665 questions that the current generation of platform reports, which classify at the topic level rather than the relational
1666 level, are not designed to answer.
1667

1668 **6.4 Demographic Patterns: What We See and What We Cannot Claim** 1669

1670 The gender-conditional patterns in our data are consistent and cross-nationally stable: men over-index on programming,
1671 finance, and religion; women over-index on health, education, and content creation. These reproduce the name-inferred
1672 patterns reported by OpenAI [28] for India, providing independent validation at the individual-conversation level.
1673

1674 The harder question is what these patterns *mean*. One reading is that gendered AI usage mirrors existing occupational
1675 and social roles: men in South Asia are over-represented in IT and finance; women carry a disproportionate share of
1676 health-care and educational responsibilities within families. Under this reading, AI usage reflects societal structure
1677 rather than reshaping it. An alternative reading, more hopeful, is that women’s disproportionate use of ChatGPT for
1678 education and coursework (particularly the striking female coursework skew we observe in India (26.1% vs. 18.2%),
1679 Pakistan (32.8% vs. 19.7%), and Nigeria (23.5% vs. 20.2%)) represents AI serving as an equalizer: a private, judgement-free
1680 educational resource that partially compensates for barriers to formal education. Our data cannot distinguish between
1681 these readings. Otis et al. [29] argue that generative AI’s impact on global inequality will depend on whether it amplifies
1682 existing advantages or compensates for existing disadvantages; the gendered usage patterns we observe are a concrete,
1683 measurable instantiation of that question, but resolving it requires larger samples and, ideally, panel data that track
1684 individual users over time.
1685

1686 The age gradient is cleaner to interpret: younger users concentrate on education and skill-building, older users
1687 on finance, professional tasks, and civic topics. This is consistent across all four countries despite their different age
1688 profiles, and it is exactly what one would expect from a life-cycle model of technology use.
1689

1690 **6.5 A Digital Sophistication Gradient** 1691

1692 The raw-metadata descriptives expose an access-and-sophistication gradient that compounds existing inequalities.
1693 Plus subscribers are concentrated almost entirely in India (116 of 134 in our sample); Nigeria has only 1 Plus user
1694 and Pakistan only 3. The behavioural gap between Free and Plus users (deeper conversations, longer prompts, more
1695 feature use) suggests that this access difference translates into qualitatively different AI experiences. At \$20/month,
1696 ChatGPT Plus costs roughly 10–20% of median monthly income in Nigeria and Pakistan; the near-zero adoption rate is
1697 unsurprising but worth stating plainly, because it means that the most capable version of the most widely used AI tool
1698 is effectively inaccessible in two of our four countries.
1699

1700 Advanced-feature adoption is modest when averaged over the full observation period: web search in 12–16% of
1701 conversations, code interpreter in 3–4%, image generation below 1%. These rates are deflated by the inclusion of early
1702 periods (2023–mid-2024) when many features were unavailable on the free tier; post-GPT-4o adoption rates are likely
1703 higher. Whether the remaining gap reflects limited awareness, limited perceived relevance, or interface friction is an
1704 open question with direct product-design implications.
1705

1706 **6.6 Limitations** 1707

1708 Several limitations qualify our findings. First, our sample is recruited through Clickworker, which introduces selection
1709 bias toward digitally literate, English-proficient, gig-economy-adjacent individuals who are not representative of
1710 the broader ChatGPT user base in each country. The prominence of *online earning* themes, for instance, is likely
1711 amplified by this channel, though the fact that the same theme does *not* appear in Nigeria (recruited through the same
1712
1713
1714
1715
1716

channel) provides some evidence that it is not purely a sampling artifact. Second, while our dataset is large by the standards of demographically grounded studies ($n = 1,252$ users, $>200K$ conversations), it is orders of magnitude smaller than server-side analyses, limiting statistical power for fine-grained subgroup comparisons. Third, some additional conversation-level labels (multitasking, human-alone, user and AI education levels) were collected only for the Indian subsample; a cross-country comparison on those dimensions awaits future work. Fourth, our topic classifiers are themselves LLM-based and may introduce systematic biases for non-English content or culturally specific topics that the classifier was not trained on. Fifth, our temporal coverage coincides with a period of rapid model improvement and feature expansion (the release of GPT-4o, reasoning models, improved tool use), making it difficult to fully disentangle user behavioural evolution from changes in system capabilities. Finally, we observe behaviour but not outcomes: we can see that users ask health questions, but we cannot assess whether the answers they receive are accurate or helpful.

6.7 Future Directions

This study is a proof of concept. The methodology (recruiting users through a crowdsourcing platform, collecting ChatGPT exports with consent, applying multiple classification pipelines on a common corpus) is straightforward to scale. A tenfold expansion of the sample, covering additional countries in Sub-Saharan Africa, Southeast Asia, and Latin America, would enable the kind of fine-grained demographic and geographic comparisons that our current sample cannot support. We see several specific directions that are feasible with this methodology but beyond the scope of the current paper:

- **High-stakes response quality.** Given the volume of health, legal, and financial queries, expert annotation of the medical, legal, and financial responses users receive on a stratified subsample would address one of the most policy-relevant questions our data raises: whether ChatGPT advice in these high-stakes domains is accurate, and whether substituting it for professional consultation improves or harms outcomes.
- **Disclosure-plus-delegation in mental-health-adjacent exchanges.** The qualitative reading in Section 5.4.1 shows that the hybrid disclosure-plus-delegation pattern is especially dense around medication, symptoms, religious identity, and relationships. Dedicated qualitative work on these exchanges, ideally with clinical input, would clarify what users are actually receiving and what the safety implications are.
- **Closer reading of women’s coursework use.** The striking female coursework skew in India, Pakistan, and Nigeria deserves dedicated qualitative work: what specific gaps is the technology filling, what does the resulting AI–student relationship look like, and does the use translate into actual learning gains or substitute for deeper engagement?
- **Longitudinal tracking.** Repeated exports from the same users over time would enable within-person analysis of how usage evolves, complementing the cross-sectional patterns we report here.
- **Multilingual analysis.** Our conversation-level language detection (Section 5.5) shows that 17–20% of conversations in India and Pakistan and over 80% in Brazil are not in English, with native-language use rising over time and concentrated in personal and emotionally expressive exchanges. Dedicated analysis of Urdu, Portuguese, Hausa, and Hindi conversations, including code-switching patterns and language-specific topic distributions, would capture dimensions of usage that our primarily English-language classification pipeline likely under-represents.
- **Causal substitution questions.** The substitution patterns we document raise causal questions our observational data cannot answer: whether users treating ChatGPT as a first-line health resource make better or worse decisions than under the counterfactual; whether religious or jurisprudential interpretation from a model trained on largely

1769 Western text affects local knowledge ecosystems over time; and whether the disclosure habit conversational
1770 AI teaches transfers to other digital systems as those systems move toward ad-supported business models that
1771 profit from intimate material.
1772

- 1773 • **Gender gaps and AI equity.** The gendered usage patterns we observe are a concrete instance of the broader
1774 question posed by Otis et al. [29]: does generative AI narrow or widen existing inequalities? A larger, more
1775 demographically balanced sample would allow this question to be addressed with statistical rigor rather than
1776 suggestive pattern-matching.
1777
- 1778 • **Community infrastructure.** We release classifier prompts, aggregate tables, and anonymized metadata to sup-
1779 port independent re-analysis. We hope this paper demonstrates the value of conversation-level, demographically
1780 grounded data on AI usage, and we encourage the research community to build on this methodology: through
1781 expanded data collection, alternative classification schemes, or cross-disciplinary analyses that connect usage
1782 patterns to outcomes in health, education, and economic welfare.
1783

1784 7 Conclusion

1785 We analyzed 202,590 ChatGPT conversations from 1,252 users across India, Nigeria, Brazil, and Pakistan, to our
1786 knowledge the first conversation-level, demographically grounded comparison of LLM usage across multiple Global
1787 South markets on a common classification pipeline. Four findings stand out: (1) users in our four countries over-
1788 index on information-seeking and writing relative to global averages, with substantial per-country heterogeneity; (2)
1789 unsupervised topic discovery surfaces culturally specific use cases (health, religion, translation, digital entrepreneurship,
1790 emotional self-reflection) that a fixed taxonomy absorbs into generic buckets; (3) *Doing* has risen steadily while a quiet
1791 upward drift in *Expressing* signals an evolving relationship between users and conversational AI; and (4) personal use
1792 is the majority everywhere, coursework rivals work in prevalence, and female coursework engagement is markedly
1793 higher in India, Pakistan, and Nigeria.
1794

1795 As we argue in Section 6, the broadest implication is that workplace productivity, the dominant lens for evaluating
1796 AI’s economic impact, captures only a minority of what these users actually do with the technology. Health advice,
1797 translation, educational support, and everyday problem-solving are forms of household production whose returns are
1798 real but invisible to standard metrics. Product design calibrated to platform-wide averages risks under-serving these
1799 use cases, and pricing that places advanced features behind a \$20/month paywall makes the most capable version of the
1800 tool effectively inaccessible in markets like Nigeria and Pakistan.
1801

1802 This study is a proof of concept: the methodology is straightforward to scale, and we release classifier prompts,
1803 aggregate tables, and anonymized metadata for replication. The questions it raises (whether AI-mediated health
1804 information helps or harms, whether gendered usage patterns reinforce or compensate for existing inequalities, what
1805 the growth of *Expressing* means for human–AI relationships) are larger than any single paper. We hope this work
1806 motivates expanded, community-driven data collection across a broader set of countries and populations.
1807

1808 References

- 1809 [1] Jenny C. Aker and Isaac M. Mbiti. 2010. Mobile Phones and Economic Development in Africa. *Journal of Economic Perspectives* 24, 3 (2010), 207–232.
- 1810 [2] Nicholas O. Alozie and Patience Akpan-Obong. 2011. The Digital Gender Divide: Confronting Obstacles to Women’s Development in Africa.
1811 *Development Policy Review* 29, 2 (2011), 161–179.
- 1812 [3] Anthropic. 2025. The Anthropic Economic Index. (2025). <https://www.anthropic.com/research/the-anthropic-economic-index>
- 1813 [4] Anthropic. 2025. *The Anthropic Economic Index: Geographic Variation in Claude Usage*. Technical Report. Anthropic. [https://www.anthropic.com/](https://www.anthropic.com/research/anthropic-economic-index)
1814 [research/anthropic-economic-index](https://www.anthropic.com/research/anthropic-economic-index)
1815
1816
1817
1818
1819
1820

- 1821 [5] Amy Antonio and David Tuffley. 2014. The Gender Digital Divide in Developing Countries. *Future Internet* 6, 4 (2014), 673–687.
- 1822 [6] Payal Arora and Aashish Rathi. 2023. The Value of AI in the Global South: Perspectives from Indian Workplaces. *AI & Society* (2023).
- 1823 [7] David Baidoo-Anu and Leticia Owusu Ansah. 2023. Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential
- 1824 Benefits of ChatGPT in Promoting Teaching and Learning. *Journal of AI* 7, 1 (2023), 52–62.
- 1825 [8] Gary S. Becker. 1965. A Theory of the Allocation of Time. *The Economic Journal* 75, 299 (1965), 493–517.
- 1826 [9] Alexander Bick, Adam Blandin, and David J. Deming. 2024. *The Rapid Adoption of Generative AI*. NBER Working Paper 32966. National Bureau of
- 1827 Economic Research. doi:10.3386/w32966
- 1828 [10] Laura Boeschoten, Theo Araujo, Jef Ausloos, Judith Möller, and Daniel L. Oberski. 2022. A framework for privacy preserving digital trace data
- 1829 collection through data donation. *Computational Communication Research* 4, 2 (2022), 388–423. doi:10.5117/CCR2022.2.002.BOES
- 1830 [11] Erik Brynjolfsson, Danielle Li, and Lindsey R. Raymond. 2023. *Generative AI at Work*. NBER Working Paper 31161. National Bureau of Economic
- 1831 Research. doi:10.3386/w31161
- 1832 [12] Aaron Chatterji, Tom Cunningham, David Deming, Zoë Hitzig, Christopher Ong, Carl Shan, and Kevin Wadman. 2025. How People Use ChatGPT.
- 1833 (2025). <https://cdn.openai.com/pdf/a253471f-8260-40c6-a2cc-aa93fe9f142e/economic-research-chatgpt-usage-paper.pdf>
- 1834 [13] Aakash Chinchure, Shail Patel, Priyanshu Jain, and Apoorv Agrawal. 2023. Demographic Patterns in AI Chatbot Usage. In *Proceedings of the AAAI*
- 1835 *Conference on Artificial Intelligence*.
- 1836 [14] Fabrizio Dell’Acqua, Edward McFowland III, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François
- 1837 Candelon, and Karim R. Lakhani. 2023. *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge*
- 1838 *Worker Productivity and Quality*. Working Paper 24-013. Harvard Business School.
- 1839 [15] Jiahong Deng and Yuanyuan Lin. 2023. The Benefits and Challenges of ChatGPT: An Overview. *Frontiers in Computing and Intelligent Systems* 2, 2
- 1840 (2023), 81–83.
- 1841 [16] Yuyang Deng, Ni Zhao, and Xin Huang. 2023. Early ChatGPT User Portrait through the Lens of Data. arXiv:2312.10078 [cs.HC]
- 1842 [17] Jonathan Donner. 2015. *After Access: Inclusion, Development, and a More Mobile Internet*. MIT Press.
- 1843 [18] Tyna Eloundou, Sam Manning, Pamela Mishkin, and Daniel Rock. 2023. GPTs Are GPTs: An Early Look at the Labor Market Impact Potential of
- 1844 Large Language Models. *arXiv preprint arXiv:2303.10130* (2023).
- 1845 [19] Cathy Mengying Fang, Sheer Karny, Chayapatr Archiwaranguprok, Yasith Samaradivakara, Pat Pataranutaporn, and Pattie Maes. 2026. AI-Wrapped:
- 1846 Participatory, Privacy-Preserving Measurement of Longitudinal LLM Use In-the-Wild. In *Proceedings of CHI*. <https://arxiv.org/abs/2602.18415>
- 1847 [20] Alison Gillwald, Onkokame Mothobi, and Broc Rademan. 2019. The State of ICT in Africa. *Research ICT Africa* (2019).
- 1848 [21] Maarten Grootendorst. 2022. BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure. *arXiv preprint arXiv:2203.05794* (2022).
- 1849 [22] Martin Hilbert. 2011. Digital Gender Divide or Technologically Empowered Women in Developing Countries? A Typical Case of Lies, Damned Lies,
- 1850 and Statistics. *Women’s Studies International Forum* 34, 6 (2011), 479–489.
- 1851 [23] Pratik Joshi, Sebastin Santy, Amar Buber, Kalika Bali, and Monojit Choudhury. 2020. The State and Fate of Linguistic Diversity and Inclusion in the
- 1852 NLP World. In *Proceedings of ACL*. 6282–6293.
- 1853 [24] Sai Keerthana Karnam, Abhisek Dash, Krishna P. Gummadi, Animesh Mukherjee, Ingmar Weber, and Savvas Zannettou. 2026. Bowling with ChatGPT:
- 1854 On the Evolving User Interactions with Conversational AI Systems. In *Proceedings of The Web Conference (WWW)*. <https://arxiv.org/abs/2602.01114>
- 1855 [25] Enkelejda Kasneci, Kathrin Sessler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan
- 1856 Günemann, Eyke Hüllermeier, et al. 2023. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. *Learning*
- 1857 *and Individual Differences* 103 (2023), 102274.
- 1858 [26] Tarek Naoos, Michael J. Ryan, Alan Ritter, and Wei Xu. 2024. Having Beer after Prayer? Measuring Cultural Bias in Large Language Models. In *ACL*.
- 1859 doi:10.18653/v1/2024.acl-long.862
- 1860 [27] Shakked Noy and Whitney Zhang. 2023. Experimental evidence on the productivity effects of generative artificial intelligence. *Science* 381, 6654
- 1861 (2023), 187–192. doi:10.1126/science.adh2586
- 1862 [28] OpenAI. 2026. *How India Uses ChatGPT: Early Evidence from OpenAI Signals*. Technical Report. OpenAI. [https://cdn.openai.com/signals/how-india-](https://cdn.openai.com/signals/how-india-uses-chatgpt.pdf)
- 1863 [uses-chatgpt.pdf](https://cdn.openai.com/signals/how-india-uses-chatgpt.pdf)
- 1864 [29] Nicholas G Otis, Solene Delecourt, Katelyn Cranney, and Rembrand Koning. 2024. *Global evidence on gender gaps and generative AI*. Harvard
- 1865 Business School Boston, MA, USA.
- 1866 [30] Siru Ouyang, Shuohang Zhang, Meiqi Zhao, Yelong Gao, and Jiawei Chen. 2023. Shifted Diffusion for Text-to-Image Generation. *arXiv preprint*
- 1867 *arXiv:2306.02115* (2023).
- 1868 [31] Jason Phang, Michael Lampe, et al. 2025. Investigating Affective Use and Emotional Well-being on ChatGPT. arXiv:2504.03888 [cs.HC]
- 1869 [32] Dejan Ravšelj, Damijana Keržič, Nina Tomažević, Lan Umek, Nejc Brezovar, et al. 2025. Higher education students’ perceptions of ChatGPT: A
- 1870 global study of early reactions. *PLOS ONE* (2025). doi:10.1371/journal.pone.0315011
- 1871 [33] Nithya Sambasivan, Shivani Kapania, Hannah Highfill, Diana Akrong, Praveen Paritosh, and Lora M. Aroyo. 2021. “Everyone Wants to Do the
- 1872 Model Work, Not the Data Work”: Data Cascades in High-Stakes AI. In *Proceedings of CHI*. 1–15.
- [34] Marita Skjuve, Asbjørn Følstad, Knut Inge Fostervold, and Petter Bae Brandtzaeg. 2021. My chatbot companion: A study of human-chatbot
- relationships. *International Journal of Human-Computer Studies* 149 (2021), 102601. doi:10.1016/j.ijhcs.2021.102601
- [35] Alex Tamkin, Miles McCain, Kunal Handa, Esin Durmus, et al. 2024. Clio: Privacy-Preserving Insights into Real-World AI Use. arXiv:2412.13678 [cs.CY]

- 1873 [36] Yan Tao, Olga Viberg, Ryan S. Baker, and René F. Kizilcec. 2024. Cultural bias and cultural alignment of large language models. *PNAS Nexus* (2024).
1874 doi:10.1093/pnasnexus/pgae346
- 1875 [37] Sherry Turkle. 2017. *Alone Together: Why We Expect More from Technology and Less from Each Other* (3rd ed.). Basic Books.
- 1876 [38] Jan A. G. M. van Dijk. 2020. *The Digital Divide*. (2020).
- 1877 [39] Mark Warschauer. 2004. *Technology and Social Inclusion: Rethinking the Digital Divide*. MIT Press.
- 1878 [40] Mark Warschauer, Wavell Tseng, Soobin Yim, Tamara Webster, Sharin Jacob, Qian Du, and Tammy Tate. 2023. The Affordances and Contradictions
1879 of AI-Generated Text for Second Language Writers. *Journal of Second Language Writing* 62 (2023), 100894.
- 1880 [41] Lixiang Yan, Lele Sha, Linxuan Zhao, Yuheng Li, Roberto Martinez-Maldonado, Guanliang Chen, Xinyu Li, Yueqiao Jin, and Dragan Gasevic. 2024.
1881 Practical and Ethical Challenges of Large Language Models in Education: A Systematic Scoping Review. *British Journal of Educational Technology* 55,
1882 1 (2024), 90–112.
- 1883 [42] Wenting Zhao, Xiang Ren, Jack Hessel, Claire Cardie, Yejin Choi, and Yuntian Deng. 2024. WildChat: 1M ChatGPT Interaction Logs in the Wild. In
1884 *Proceedings of ICLR*. <https://arxiv.org/abs/2405.01470>
- 1885 [43] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P. Xing, Hao
1886 Zhang, Joseph E. Gonzalez, and Ion Stoica. 2023. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. In *Advances in Neural Information
1887 Processing Systems (NeurIPS)*.

1888 A BERTopic Pipeline Details

1889 This appendix documents the full unsupervised topic-modelling pipeline summarized in Section 4.

1890
1891
1892 *Embedding.* Each conversation is represented by a dense semantic embedding obtained from Google’s `gemini-embedding-001`
1893 model, which produces 3,072-dimensional vectors. The embedding input is the first ten user–assistant turns of the
1894 conversation, serialized in the `[User]: ... \n[Assistant]: ...` format, truncated at a per-message cap of 5,000
1895 characters; inputs exceeding the model’s token limit are truncated by the API. Using the first ten turns rather than
1896 the whole conversation keeps per-conversation cost bounded and, in practice, captures the topic of the conversation
1897 well (users typically establish their task in the opening exchanges). All embedding vectors are L2-normalized before
1898 clustering.
1899

1900
1901
1902 *Initial clustering.* For each country we fit MiniBatch K -means with $k = 500$ on the embedding matrix. The deliberately
1903 high k produces narrow, internally coherent micro-clusters (e.g. “debugging Python code for web scraping” rather than
1904 the generic “programming”). Starting from an over-segmented solution lets the subsequent aggregation be driven by
1905 semantic similarity rather than forced early generalization.
1906

1907
1908 *Hierarchical agglomeration.* The 500 K -means centroids are then merged with agglomerative hierarchical clustering
1909 using cosine distance. The cutoff is selected per country by visual inspection of the dendrogram for natural breakpoints,
1910 targeting approximately 50 interpretable top-level topics, a number that balances cross-country comparability with
1911 sufficient granularity to reveal local patterns.
1912

1913
1914 *Topic labelling and refinement.* Each resulting cluster is labelled by summarization with `gpt-4o-mini`, using the top
1915 documents and most distinctive keywords in the cluster as context. To enhance label validity, we used Claude Sonnet
1916 4.6 as a judge: clusters with ambiguous or low-confidence labels were either relabelled or flagged as non-homogeneous.
1917 Conversations in flagged clusters were then reassigned to defined topics by cosine-similarity scoring against cluster
1918 centroids, using a per-cluster z -score threshold of 1.5 as the reassignment floor. The final pipeline yields **50, 45, 36,**
1919 **and 53** topics for India, Nigeria, Brazil, and Pakistan respectively, covering 91–95% of conversations per country; the
1920 remaining unassigned conversations (typically very short, off-topic, or multilingual/code-heavy) are excluded from
1921 BERTopic analyses but retained elsewhere.
1922
1923

Cross-country theme grouping. For the ten-theme cross-country view in Section 5.2, we group the per-country clusters into ten broad themes (Programming/Tech, Finance/Earning, Writing/Creative, Religion, Translation/Language, Health/Medical, Education/Academic, Job/Career, Content Creation, Current Affairs) by keyword matching on the cluster labels. This is a lossy aggregation (some cluster labels match multiple themes) but it makes the four countries comparable on a shared vocabulary while retaining the per-country cluster resolution for the cluster-level figures.

B Classifier validation

To assess the reliability of the two LLM-based conversation-level classifiers—*Work/Coursework/Personal* (Section 5.3) and *Asking/Doing/Expressing* (Section 5.4)—we hand-labeled a stratified random sample of 50 conversations per country (194 valid items pooled across India, Brazil, Pakistan, and Nigeria after dropping ambiguous cases). One of the authors, blind to the model’s prediction, assigned a gold label using the same definitions given to the classifier prompt. We then computed pooled accuracy, macro-F1, and Cohen’s κ against the model labels (Table 5).

Both classifiers reach substantial pooled agreement (Table 5). The *Work/Coursework/Personal* classifier is reliable: pooled accuracy is 0.88 and $\kappa = 0.82$ (substantial agreement), with per-country accuracy ranging from 0.84 (Nigeria) to 0.94 (India). The *Coursework* class is particularly clean (per-class F1 ≈ 0.95); the dominant error mode is the model labeling some *Personal* conversations as *Work* (15/194 pooled cases), reflecting genuine overlap when personal queries borrow professional/work-style language.

For the *Asking/Doing/Expressing* classifier, the *Asking* and *Doing* classes are individually reliable (F1 = 0.84 and 0.77), but *Expressing* has high recall (0.78) and low precision (0.35): the classifier frequently labels conversations as *Expressing* when the gold label is *Asking* or *Doing*. This is consistent with the qualitative finding in Section 5.4.1: very few *Expressing* conversations in our corpus are purely affective, and most interleave a personal disclosure or emotional framing with a concrete request for information or task delegation. The classifier’s recall on the human-gold *Expressing* class is therefore high; the precision drop reflects the same hybrid character, where conversations the annotators called *Asking* or *Doing* also contained enough self-disclosure for the model to predict *Expressing*. We treat the classifier’s *Expressing* prevalence as a noisy upper bound and report *Expressing*-conditioned results with that caveat.

Table 5. Pooled validation metrics against human gold labels on a stratified sample of ~ 50 conversations per country. Both tasks reach substantial agreement; the residual error in *Asking/Doing/Expressing* is driven mainly by over-prediction of *Expressing* (see text).

Task	Acc. (95% CI)	Macro-F1	κ
Work/Coursework/Personal	0.88 (0.83–0.92)	0.88	0.82
Asking/Doing/Expressing	0.81 (0.75–0.86)	0.70	0.55

C Cross-Dataset Validation: WildChat

To assess the generalizability of our findings beyond our recruited cohort and to compare with global trends, we analyze conversation topics in WildChat-4.8M [42], an independent corpus of ChatGPT interactions collected via a Gradio-based research proxy. WildChat provides geographic metadata (derived from IP addresses) for a large, diverse user base. From the top 20 countries by volume we situate our India and Brazil findings within a broader global comparison.

Table 6. Per-country Cohen’s κ for each classifier.

Country	n	W/C/P κ	A/D/E κ
India	47	0.90	0.51
Brazil	50	0.79	0.72
Pakistan	48	0.84	0.53
Nigeria	49	0.76	0.44
Pooled	194	0.82	0.55

Table 7. WildChat sample overview. Top 20 countries by total conversation volume, before and after sampling users with 10+ conversations. For United States, Russia, China, Germany, and Hong Kong we further sampled 500 users each.

Country	Total Conversations	Sampled Users	Sampled Conversations
United States	944,046	500	36,394
Russia	262,050	500	36,770
China	253,752	500	14,296
Germany	114,837	500	34,507
United Kingdom	103,614	445	43,936
Japan	98,810	446	12,924
Vietnam	75,220	258	59,186
Hong Kong	70,225	500	14,526
India	62,804	577	17,315
France	62,309	483	19,430
South Korea	61,894	231	15,697
Brazil	57,702	427	16,033
Canada	51,265	320	18,228
Taiwan	41,691	409	20,256
Australia	37,080	259	12,707
Italy	36,827	262	9,177
Spain	33,019	240	10,178
The Netherlands	32,578	207	9,048
Singapore	30,899	375	11,567
Egypt	27,708	219	4,640

C.1 Sampling and Classification

We selected the top 20 countries by total conversation volume in WildChat (Table 7). To ensure sufficient per-user coverage, we sampled users with 10 or more conversations; for five high-volume countries (United States, Russia, China, Germany, and Hong Kong), we further capped the sample at 500 users each to keep classification costs tractable. India and Brazil, as primary focus countries of our study, were retained without capping. This yielded 398,050 classified conversations across 20 countries. We applied the same OpenAI taxonomy classifier (GPT-4o, 24 fine-grained categories, 7 coarse domains) used for our primary dataset, enabling direct comparison.

A key difference between the two datasets is their user populations. WildChat users self-selected into a research proxy to access ChatGPT for free, creating a sample that skews toward technically sophisticated, often English-speaking users, particularly in non-Anglophone countries. Our primary dataset, by contrast, represents a broader demographic cross-section through crowdsourced recruitment.

C.2 Coarse Topic Distribution Across Countries

Figure 25 presents the coarse topic composition across all 20 WildChat countries. Several patterns are immediately apparent. *Writing* dominates in Anglophone countries, reaching 66.0% in the United Kingdom and 46.3% in the United States, driven primarily by fiction writing, which alone accounts for 40.2% and 18.5% of conversations in these countries respectively. *Multimedia* (primarily image generation) is concentrated in East Asian countries: Japan (49.8%), Hong Kong (43.2%), and China (41.9%), reflecting the popularity of AI-generated art and character illustration in these markets. *Technical Help* varies from under 1% (Vietnam) to 37.3% (India), the single largest source of cross-country variation.

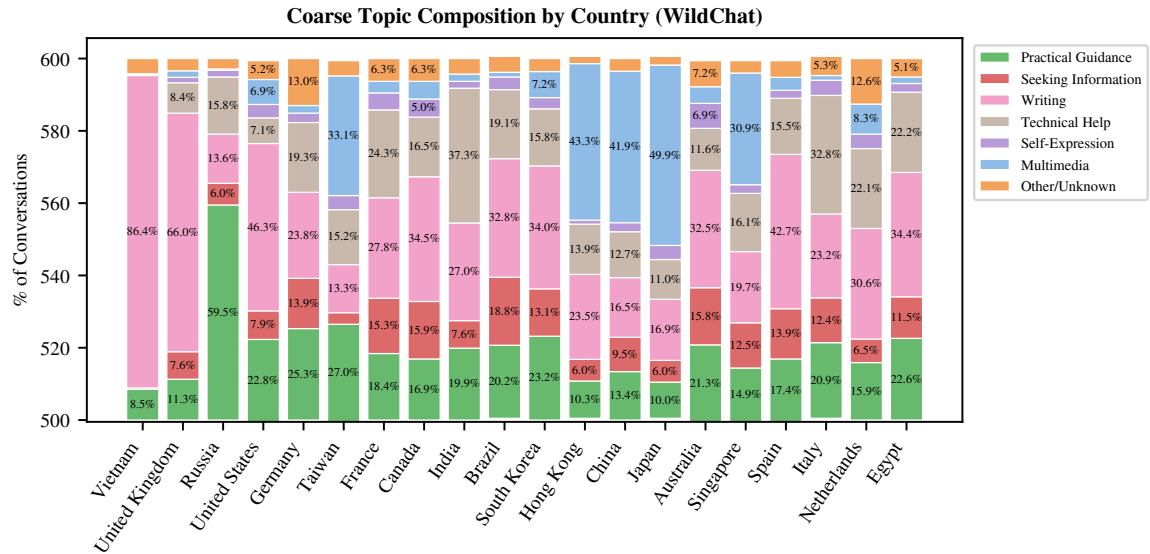


Fig. 25. Coarse topic composition across 20 WildChat countries, sorted by total conversations.

C.3 India and Brazil in Global Context

Figure 26 shows the coarse topic profiles of India and Brazil alongside four high-volume comparison countries (United States, United Kingdom, Germany, France), and Figure 27 quantifies their deviation from the 20-country mean.

India in WildChat is strikingly dominated by *Technical Help* (37.3%, +20.5pp above the 20-country mean), driven almost entirely by computer programming at 34.4% of all Indian conversations, ranking first among all 20 countries by a wide margin. This reflects the composition of Indian WildChat users: technically proficient individuals who discovered and adopted a research proxy interface, likely drawn disproportionately from India's software engineering workforce. *Writing* is correspondingly lower (−5.3pp), and *Multimedia* usage is minimal (2.0%, −10.4pp). *Edit or critique of provided text* ranks second (13.0%, rank 3/20).

Brazil presents a markedly different WildChat profile. *Seeking Information* is the most over-represented category (+8.6pp above the 20-country mean), with *specific information* queries constituting 18.0% of conversations, the highest share among all 20 countries. *Computer programming* is substantial but less dominant (17.3%, rank 6/20). *Translation* is notably elevated (6.4%, rank 6/20), reflecting Portuguese–English bilingual needs.

2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132

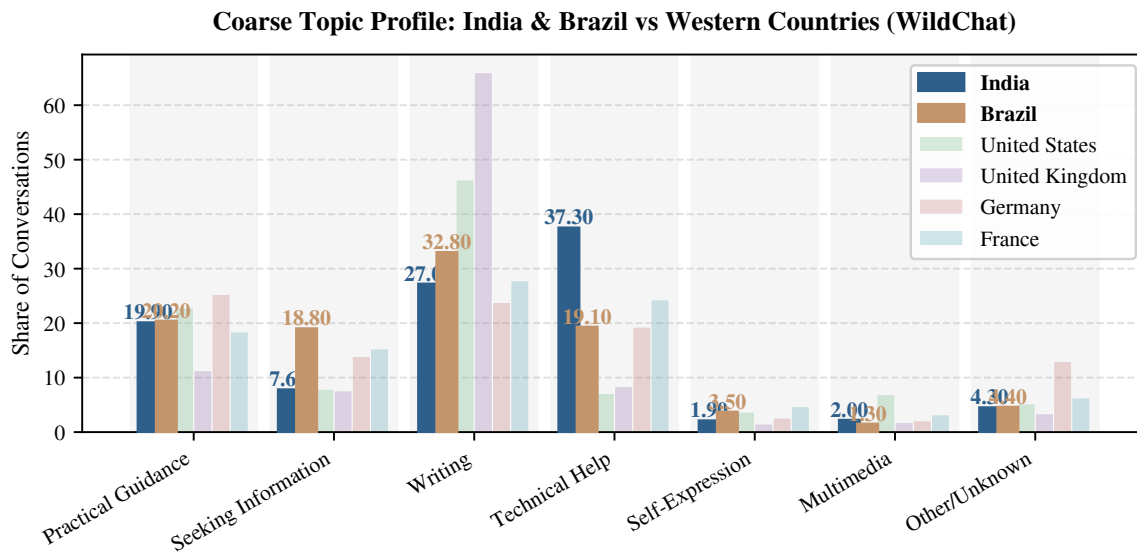


Fig. 26. Coarse topic profiles of India and Brazil compared to four high-volume comparison countries (WildChat).

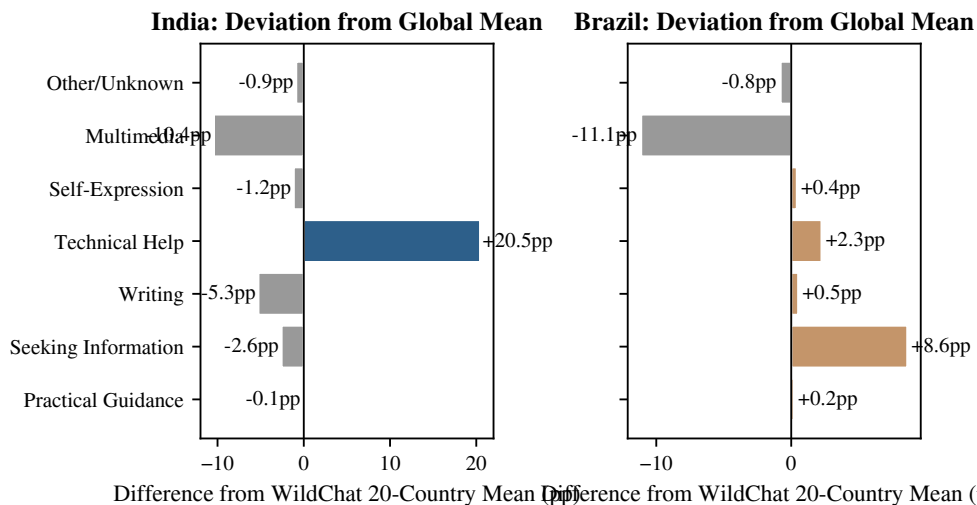


Fig. 27. Deviation of India (left) and Brazil (right) from the WildChat 20-country mean in coarse topic share (percentage points).

C.4 Comparing WildChat with Our Dataset

Figure 28 compares the WildChat topic distributions for India and Brazil with our ChatGPT-export data.

Divergences. The most striking difference is the magnitude of computer programming: 34.4% in WildChat India vs. 7.6% in our data, and 17.3% in WildChat Brazil vs. 2.4% in ours. This 4–7× inflation directly reflects WildChat’s selection bias toward technically sophisticated users who navigate a research proxy interface. Conversely, our data shows substantially higher shares of *specific information* seeking for India (19.6% vs. 7.0%) and comparable shares for

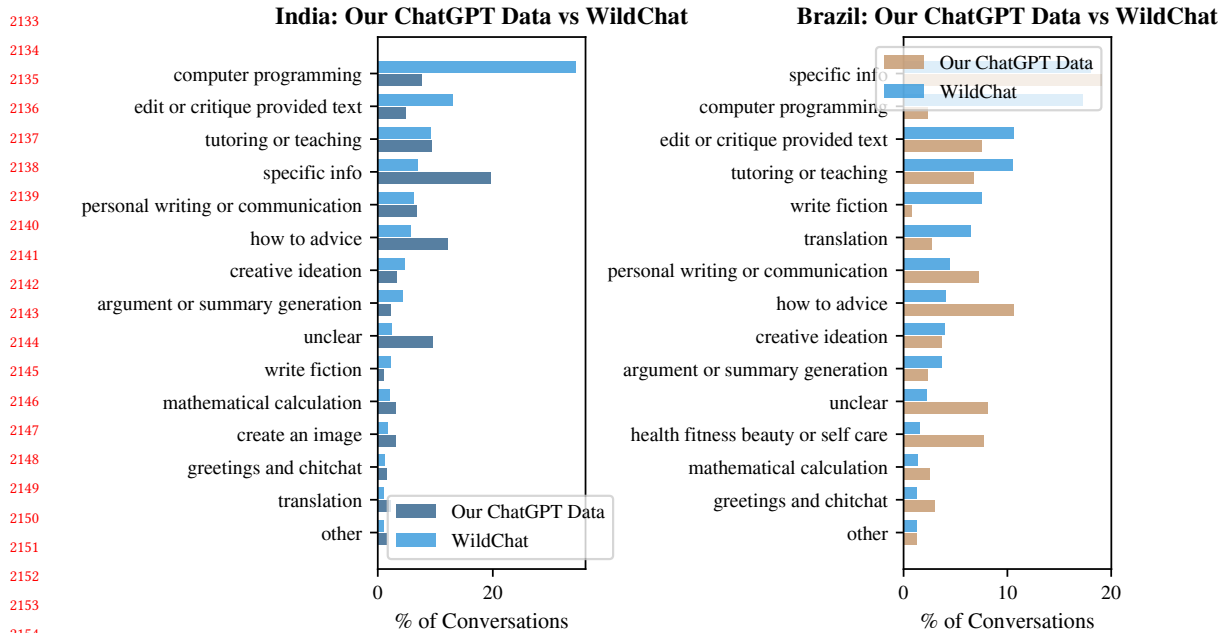


Fig. 28. Fine-grained topic comparison between our ChatGPT-export data and WildChat for India (left) and Brazil (right).

Brazil (19.1% vs. 18.0%), along with higher shares of *how-to advice* and *health/fitness/beauty*, categories that characterize broader, non-technical, personal conversations. These divergences validate a methodological concern: proxy-based corpora like WildChat systematically over-represent technical users and under-represent the information-seeking and practical guidance use cases that dominate among general populations.

Convergences. Despite these differences, several patterns are robust across both datasets. First, India leads in computer programming share in both WildChat (rank 1/20) and our dataset (7.6%, highest among our four countries), consistent with India's IT-sector-driven usage profile regardless of sampling frame. Second, Brazil consistently shows higher engagement with translation tasks (6.4% in WildChat vs. 2.8% in our data). Third, creative fiction writing is comparatively low in both India and Brazil across both datasets (2.3% in WildChat India, even smaller in ours) compared to 18.5% for the US in WildChat, suggesting that creative-fiction use varies substantially by country and language context.

Implications. The comparison underscores that topic distributions in user-AI interaction studies are sensitive to collection methodology and the resulting user sample. WildChat's proxy collection yields a portrait skewed toward users who actively sought out a research frontend, while our recruited cohort captures a broader spectrum of use cases. The convergence of robust findings across these methodologically distinct datasets (India's programming emphasis and the lower salience of creative fiction in both) strengthens confidence that those patterns reflect genuine occupational and linguistic differences rather than artifacts of either method.

D Additional Figures

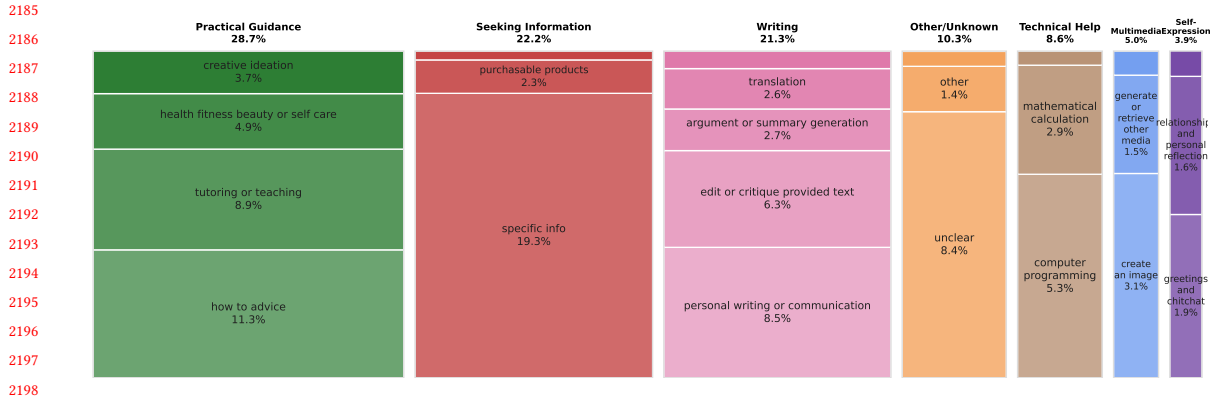


Fig. 29. Treemap of conversation topics across all four countries under the OpenAI taxonomy (Section 5.1). Outer tiles are coarse domains; inner tiles are fine-grained categories.

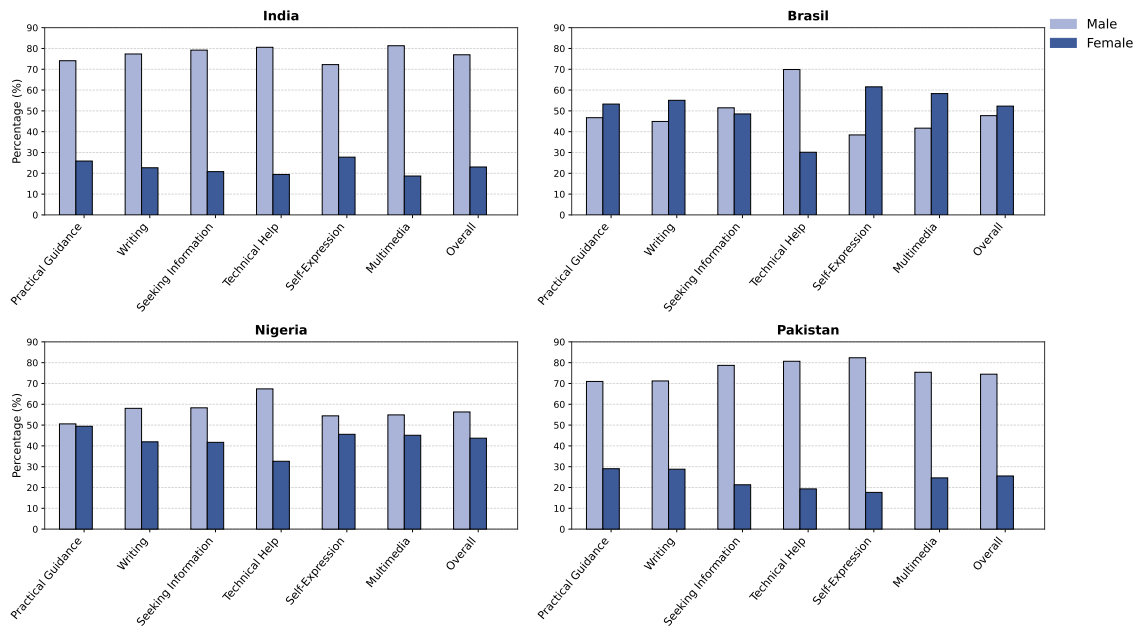


Fig. 30. Share of conversations by gender for each coarse topic, by country (Section 5.1). The “Overall” bar provides the baseline gender split; deviations from it indicate topic-level over- or under-representation.

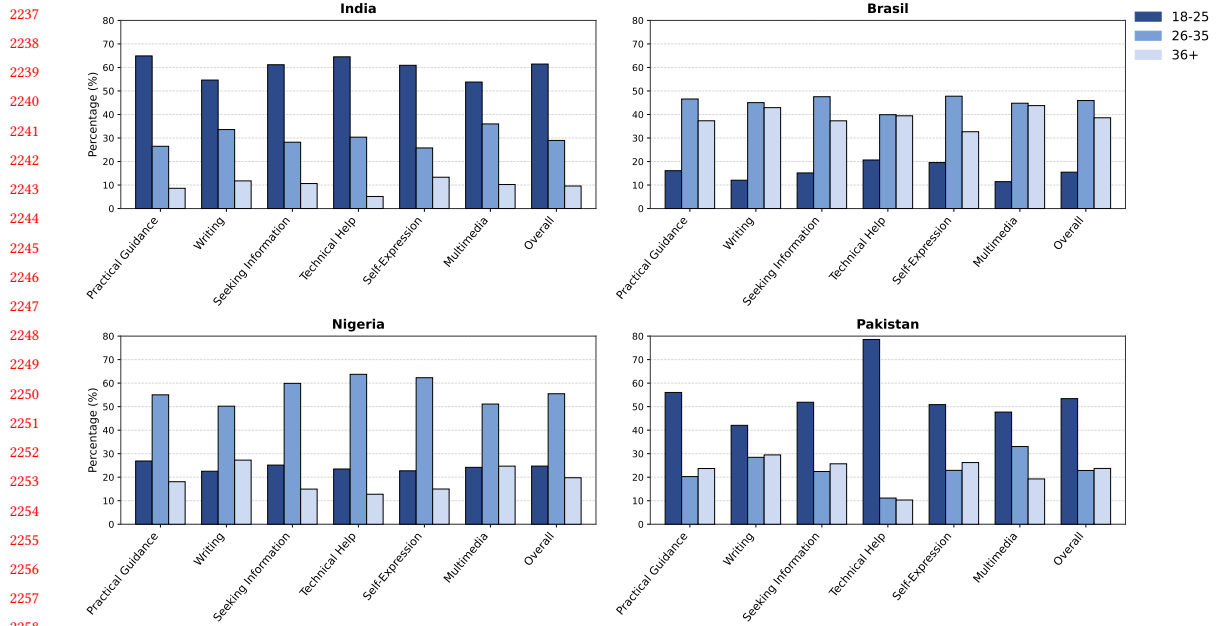


Fig. 31. Share of conversations by age group for each course topic, by country (Section 5.1). The “Overall” bar provides the baseline age distribution.

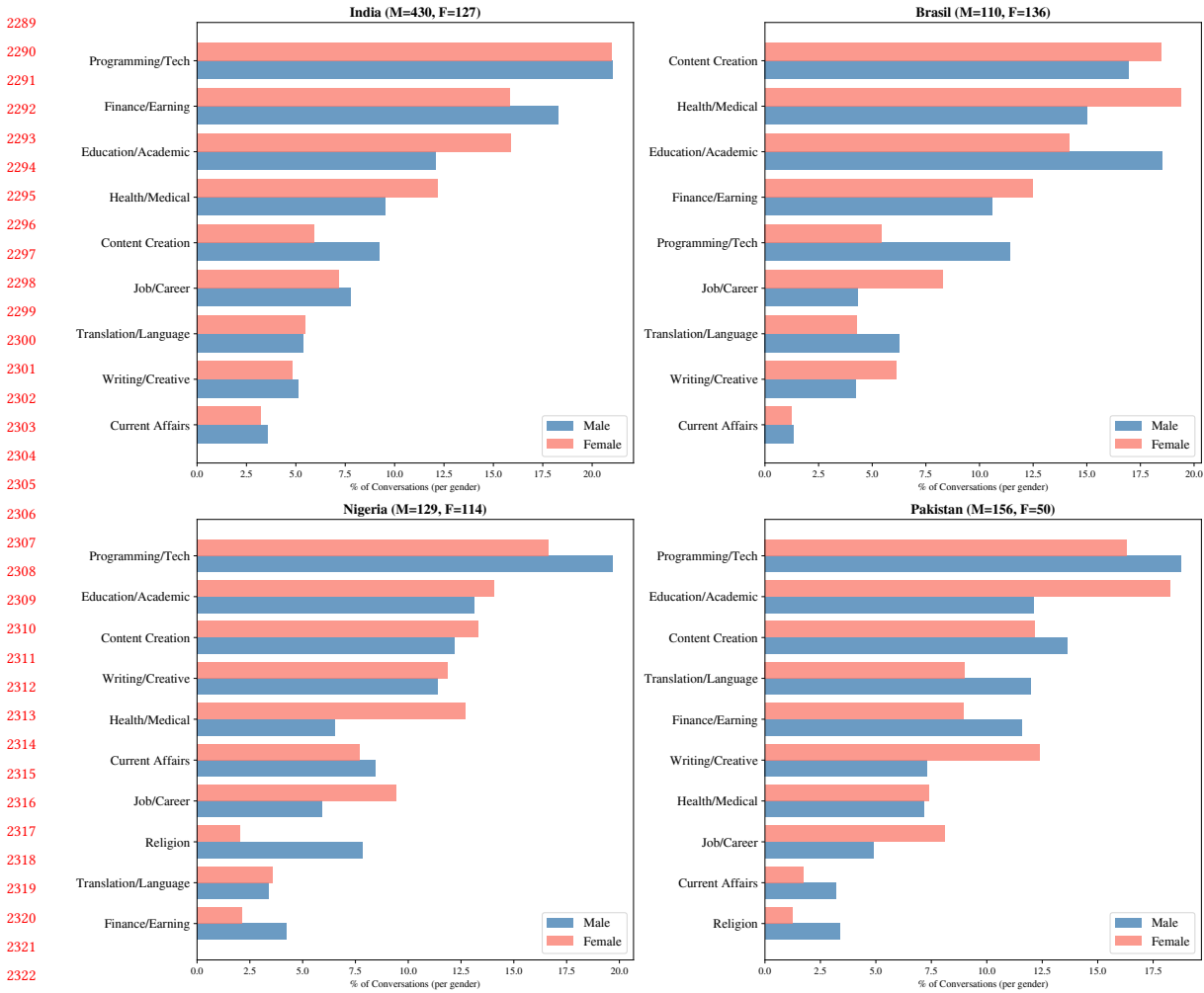


Fig. 32. Unsupervised themes by gender across countries (Section 5.2). Each bar shows the share of a gender’s conversations assigned to a theme. Male over-representation in programming, finance, and religion; female over-representation in health, content creation, and education.

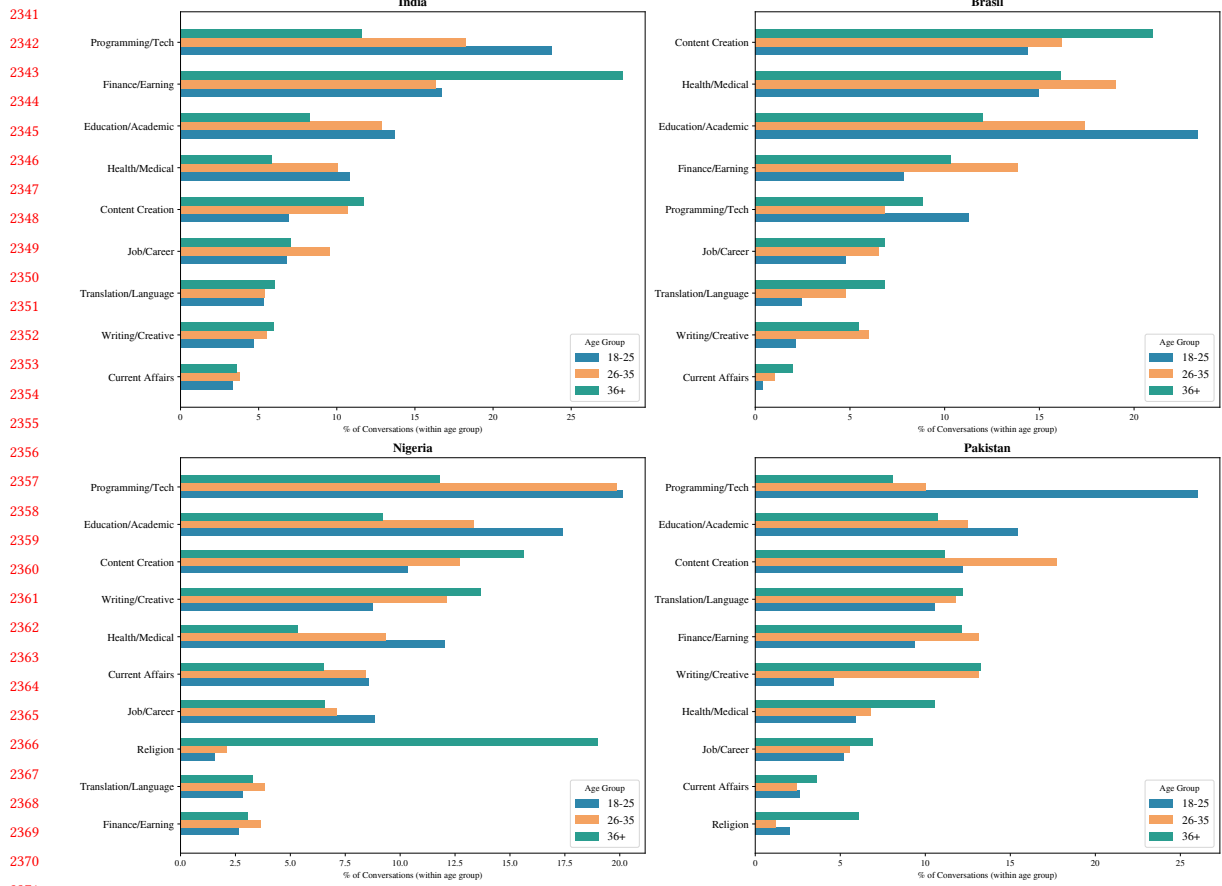


Fig. 33. Unsupervised themes by age group across countries (Section 5.2). Younger users concentrate on education and programming; older users shift toward finance, content creation, and religion.

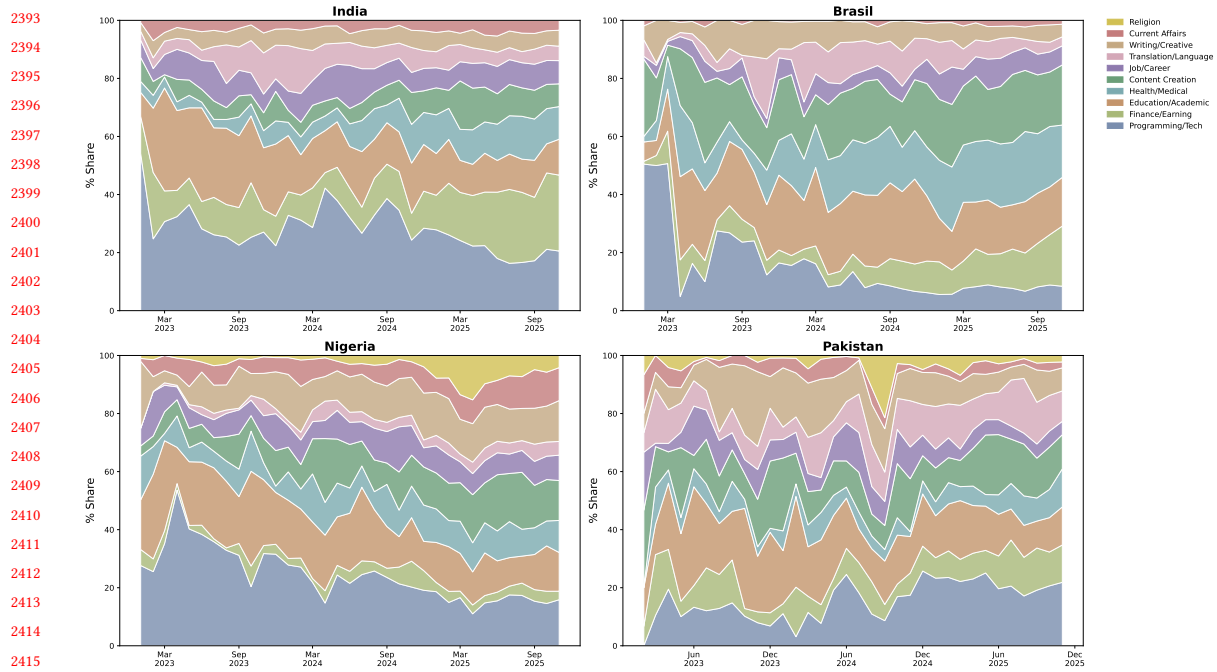


Fig. 34. Temporal shift in unsupervised theme share (stacked area) by country (Section 5.2).

2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444

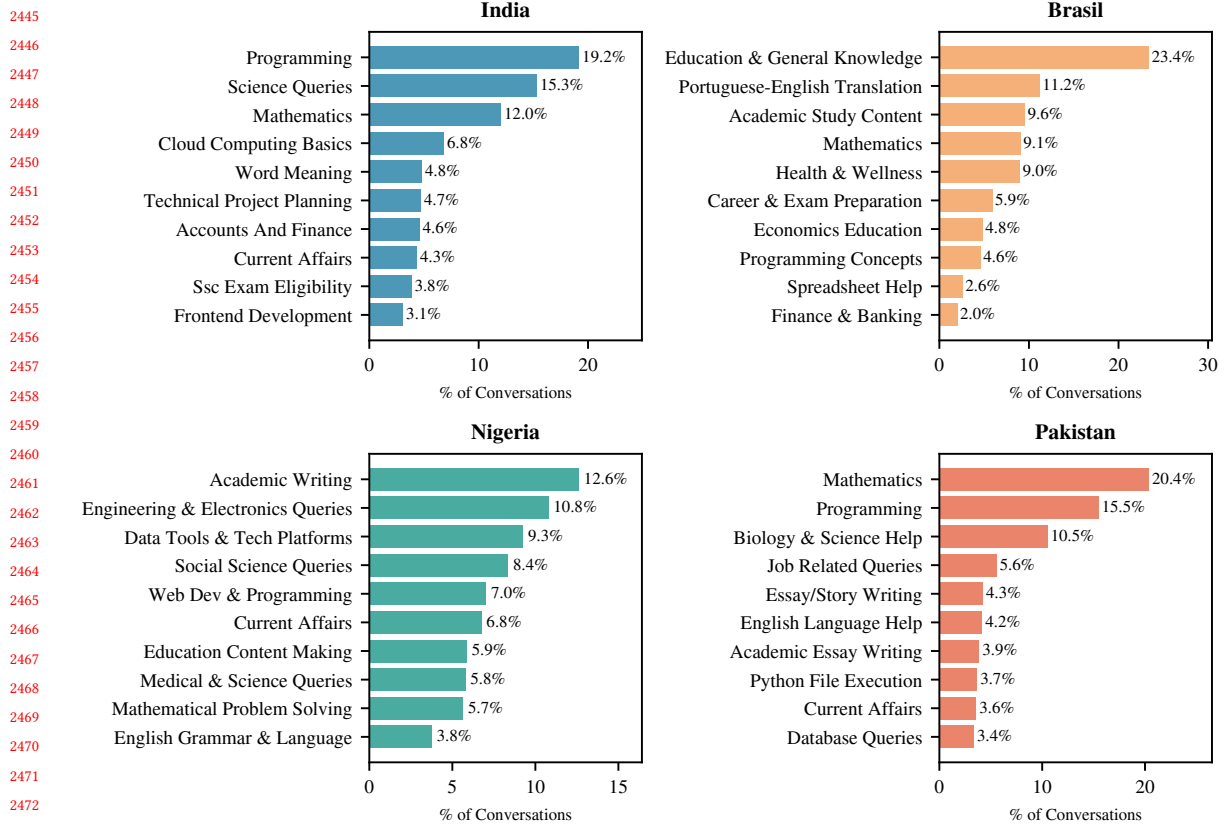


Fig. 35. Top 10 unsupervised topic clusters for coursework conversations, by country (Section 5.3).

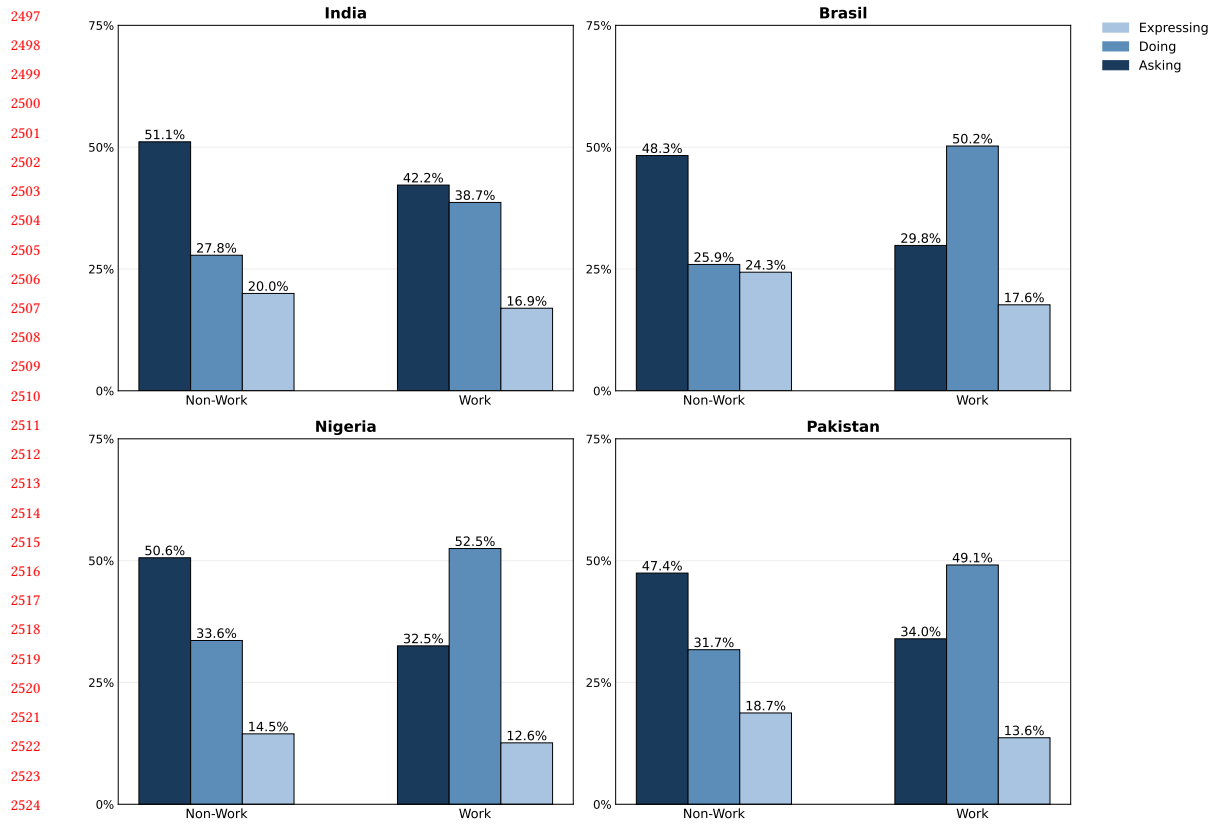


Fig. 36. Share of Asking/Doing/Expressing conversations by task purpose (work vs. non-work) and country (Section 5.4).

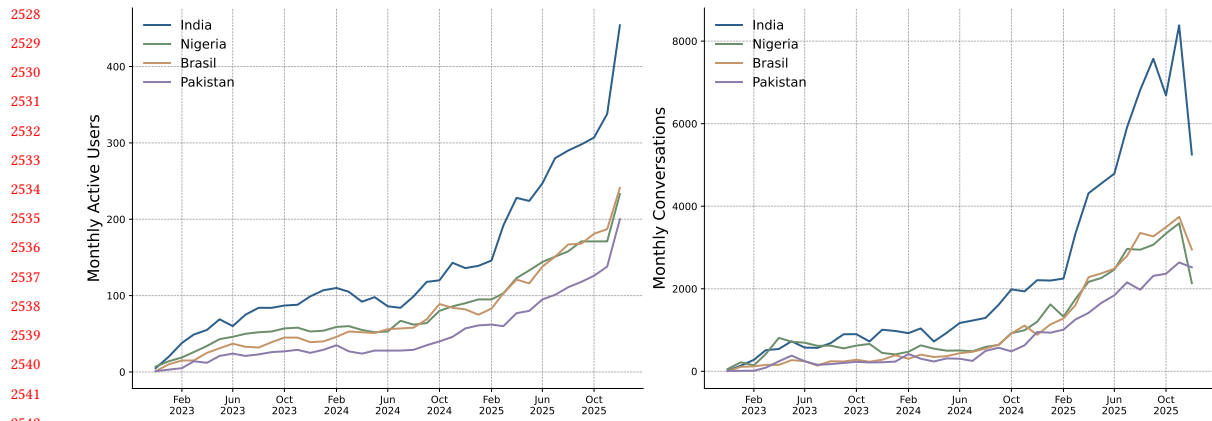


Fig. 37. Monthly active users (left) and monthly conversations (right) by country (Dec 2022–Feb 2026). See Section 3.

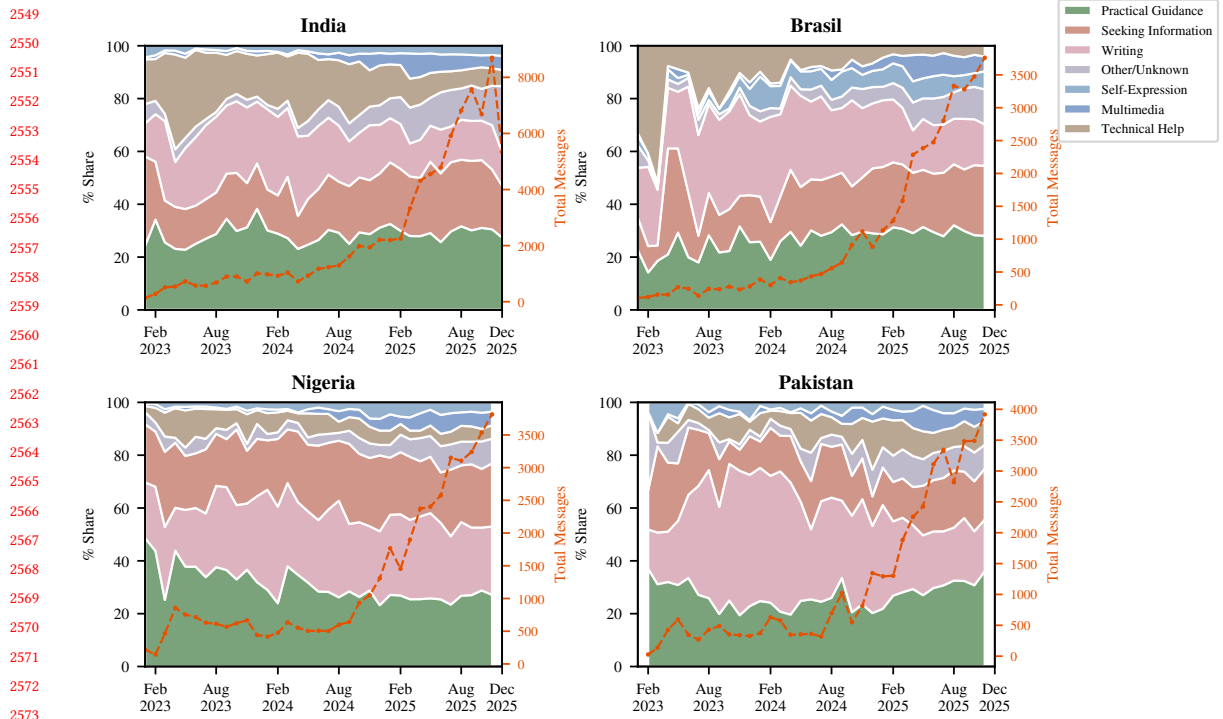


Fig. 38. Monthly coarse-topic share (stacked area, left axis) and total conversation volume (dashed line, right axis) by country, under the OpenAI taxonomy. Deferred from Section 5.1: the in-text discussion uses the line-plot version (Figure 5); this stacked-area rendering is included here for readers who prefer to see volume and share on the same chart.