



From Validation to Discovery: An Inverse-Docking Experiment for Culturally Calibrated Synthetic Personas Across Five Geographies and Two Population Types

Uday Wagh ·

CITE AS
[openxiv:cs.AI.2026.00001](https://openxiv.org/abs/cs.AI.2026.00001)
ISSN
3120-9556 (online)
LICENSE
arXiv-nonexclusive-distrib

POSTED
2026-05-27
VERSION
v1
SUBJECT
cs.AI (+1 cross-listed)

AI DISCLOSURE

COAUTHOR

COVER EVIDENCE

TRANSPARENCY · partial

IDENTITY · partial

PROVENANCE · strong

CITATIONS · strong

MATH · partial

INTEGRITY · partial

CANONICAL RECORD

<https://openxiv.net/abs/cs.AI.2026.00001>

Cite as: openxiv:cs.AI.2026.00001

Live verification record is maintained on the canonical abstract page.

DOI will be deposited and back-filled once Crossref membership clears.



scan to open

From Validation to Discovery: An Inverse-Docking Experiment for Culturally Calibrated Synthetic Personas Across Five Geographies and Two Population Types

Uday Wagh
TygrX Inc. / TwinSim
uday@twinsim.ai

Draft v1.0 – May 2026

Abstract

Synthetic persona platforms are commonly used as instruments for testing existing concepts against simulated panels. We report an inverse experiment: open-ended pain elicitation from culturally calibrated synthetic personas, followed by symmetric validation against external venture-market evidence in each persona’s market. We ran five studies across India, the United Arab Emirates, Australia, Southeast Asia, and Germany, covering two population types: B2C consumers and B2B finance and compliance professionals. In total, 1,433 personas produced 212 distinct pain themes. Between 40% and 79% of high-volume themes mapped to currently funded local startups or post-cutoff category-forming activity; the remaining 21% to 60% were classified as partial-gap or unowned commercial space. Validation rates varied with market context: India B2C returned 79%, Germany B2B returned 58%, and UAE, Australia, and Southeast Asia returned 40–43%. In the Southeast Asian mixed-country study, themes self-stratified by country, including Filipino-heavy remittance and motorcycle-taxi themes, Malaysian-heavy prayer and Ramadan themes, and Thai-heavy banana-leaf and motorbike themes. Across all five studies, personas repeatedly elevated pains where funded incumbents addressed an adjacent problem layer rather than the persona-named friction itself. We define a Discovery Index for measuring the share of persona-surfaced high-volume themes already matched by funded venture activity. The results suggest a distinct discovery use case for synthetic personas, separate from the dominant stimulus-to-response validation paradigm, while also identifying clear limits requiring independent replication.

1 Introduction

Large language models (LLMs) have made it possible to construct synthetic agents and synthetic respondents at low cost. Prior work has studied generative agents that simulate believable behavior over time [1], language-model conditioned samples that approximate response distribu-

tions of human subgroups [2], and LLM-based replications of human-subject experiments [3]. More recent work has evaluated whether synthetic personas can dock against known human interview protocols, especially in startup validation contexts [4].

The dominant applied use case, however, remains stimulus-to-response validation. A team brings an idea, product, message, prototype, or interview protocol to a synthetic panel. The synthetic respondents react. The output is then interpreted as a proxy for how real users might respond, often with a claim of speed, cost reduction, or directional parity to human research. This framing is also the target of the strongest skeptical literature. Critics argue that synthetic users can produce research theater, overconfident mimicry, demographic stereotyping, positivity bias, and findings that borrow the authority of user research without the accountability of real human data [6, 7, 8, 9].

This paper tests a different use case. Instead of bringing a stimulus to a synthetic panel, we ask whether a culturally grounded panel can surface the question space itself. In other words, can synthetic personas, when asked open-ended discovery questions from within culturally specific life contexts, produce pain themes that correspond to real venture activity in the markets they represent? We call this an inverse-docking experiment because the docking target is not a known human response dataset. The docking target is an external venture market: funded local startups, category-forming post-cutoff startups, adjacent funded categories, and unowned spaces.

The research question is:

Do culturally grounded synthetic personas, when asked open-ended discovery questions, produce pain themes that map to real venture-market activity in their represented geographies and population types?

A positive answer would not prove that synthetic personas replace human research. It would establish a narrower claim: synthetic personas may be useful as a hypothesis-generation instrument when their outputs are

evaluated symmetrically against external evidence rather than accepted at face value.

We make four contributions. First, we introduce an inverse-docking methodology for synthetic-persona discovery. Second, we report a five-study dataset spanning four B2C consumer geographies and one B2B professional geography, totaling 1,433 personas. Third, we define a Discovery Index, the proportion of persona-surfaced high-volume themes that map to currently funded or category-forming venture activity in the relevant market. Fourth, we report a repeated structural pattern: personas surfaced adjacent pain layers inside already-funded categories, suggesting a discovery signal that is different from direct retrieval of known startups.

2 Related Work

2.1 Synthetic agents and silicon samples

Park et al. [1] introduced generative agents as computational agents that use LLMs, memory, reflection, and planning to produce believable individual and social behavior. Argyle et al. [2] argued that LLMs can be conditioned on socio-demographic backstories to simulate human samples, coining the term algorithmic fidelity for the ability of a model to reproduce subgroup response patterns. Aher et al. [3] proposed Turing Experiments for evaluating how LLMs replicate findings from human-subject studies.

These works share a core assumption relevant to the present study: conditioning matters. LLM outputs are not merely one undifferentiated model voice. They can vary meaningfully with context, prompt, agent memory, or demographic framing. The present work extends this assumption into commercial discovery. We ask not whether synthetic personas reproduce known survey results or known interview themes, but whether their open-ended pain statements map to a changing external market.

2.2 Docking synthetic personas against human data

The closest published prior art is Teutloff’s study of synthetic founders [4]. That work docks human-subject founder interviews against synthetic founder and investor personas using the same interview protocol. It reports convergent, partial, human-only, and synthetic-only themes. The method is important because it treats synthetic output as something to be compared against an external reference rather than accepted directly.

The present work is inverse to that design. We do not dock synthetic personas to a known human response dataset. We elicit open-ended pains first, classify them into locked taxonomies, and then validate the resulting themes against venture-market evidence. Where Teutloff measures fidelity to known responses, we measure discovery against an external commercial environment.

2.3 Skeptical literature on synthetic users

The skeptical position is also central to this paper. NN/g defines synthetic users as AI-generated profiles that attempt to mimic user groups and warns that user research needs real users for most evaluative decisions [6]. ACM Interactions critiques synthetic personas as a potential fallacy in which LLM completions are treated as evidence without human-grounded validation [7]. MeasuringU reviews experiments with synthetic users and points to broader validity concerns [8]. A systematic review of 182 papers on synthetic participants similarly raises concerns about cognitive misalignment, stereotyping, and limits of behavioral simulation [9].

Our design accepts much of this critique. We do not claim that synthetic personas are human subjects, have agency, or can validate demand. We use them as hypothesis generators and require external validation. The paper’s key methodological choice is therefore symmetric validation: every theme that crosses the high-volume threshold is checked against external market data and assigned a status.

3 Method

3.1 Persona infrastructure

Each study used culturally stratified synthetic personas generated inside the TwinSim platform. Each persona was generated from a 16-section cultural seed profile and constrained by a voice-discipline layer intended to prevent register inflation and generic LLM voice. The full seed-generation method and the voice-discipline mechanism are not disclosed in this paper. This omission is deliberate: the paper discloses the experiment, prompts, classification method, validation status definitions, and aggregate results, while reserving the persona construction mechanism for a separate technical report.

The India persona set had previously been externally calibrated against Indian public data sources with a Spearman correlation of 0.839 across selected behavioral and demographic variables. Equivalent calibration work was conducted for the other markets, but the detailed calibration schemas are outside the scope of this paper. This paper should therefore be read as an empirical report on a discovery instrument, not as a full technical disclosure of the persona generator.

The underlying model was DeepSeek V3 via OpenRouter, with Claude Haiku used as fallback for failed generations. DeepSeek V3 is a mixture-of-experts LLM described in the DeepSeek-V3 technical report [5]. The internal study logs treat July 2024 as the model knowledge cutoff used for the post-cutoff validation argument. The studies themselves ran between March and May 2026.

3.2 Discovery prompts

The B2C prompt was issued to 1,258 personas across India, UAE, Australia, and Southeast Asia:

Given your life, which problem have you faced where the problem is not just obvious, it is so common that it is not even considered a problem to be solved? Avoid extremely common problems like better air conditioner. Things like cracked heels or women in India are true unique problems, take that as example and answer.

The B2B prompt was issued to 175 German finance and compliance personas:

Given your professional role, name a problem so common in your daily work that nobody considers it solvable. Avoid generic complaints about software or meetings. Think about the friction you have stopped noticing because it has always been there.

The prompt deliberately avoided product stimuli, examples of possible startups, or solution suggestions. The B2C prompt contained a culturally concrete seed example to push respondents away from generic consumer complaints. The Germany B2B prompt removed the cracked-heels example and instead specified professional friction.

3.3 Studies and samples

Five studies were run between March and May 2026. Table 1 summarizes the sample composition.

3.4 Classification pipeline

Persona responses were classified using a three-stage hybrid pipeline.

1. **Regex pass.** Theme-specific keyword patterns were defined in English and relevant local languages or romanizations: Hinglish for India, Bahasa Malay and Bahasa Indonesia where relevant, Tagalog, Thai, Arabic romanization for UAE, and German for Germany.
2. **TF-IDF similarity pass.** Responses not captured cleanly by regex were compared to multilingual theme-anchor descriptions using TF-IDF cosine similarity.
3. **Manual unmatched pass.** Unmatched responses were reviewed manually to identify net-new themes. When a new theme was accepted, it was added to the taxonomy and the pipeline was rerun.

After iterative development, each geography’s taxonomy was frozen as Locked Taxonomies v1.0. Future studies can therefore measure new samples against fixed theme definitions. Coverage ranged from 66.0% to 94.2%, meaning that 66.0% to 94.2% of responses matched at least one locked theme. The unmatched tail was not forced into categories.

TF-IDF was used instead of multilingual sentence-transformer embeddings because of local environment constraints during analysis. This is a limitation. It likely reduces recall for semantically equivalent responses expressed with different local phrasing. A replication should rerun the classification layer with modern multilingual embeddings while preserving the locked taxonomy for comparability.

3.5 Symmetric validation

For each high-volume theme, defined as five or more personas except in UAE where the smaller sample used a threshold of four, the author searched for currently funded local startups or commercial actors addressing that exact problem in the relevant market. Validation searches used public funding news, startup databases, company directories, and direct company-site verification. The validation window was March to May 2026.

Each high-volume theme was assigned one of four statuses:

- **Validated:** multiple funded local competitors or established commercial actors exist for the theme.
- **Category-forming:** funded local competitors raised meaningful capital or launched after the model cutoff, making direct training-data retrieval less plausible.
- **Partial gap:** adjacent products exist, but no funded actor clearly owns the specific persona-named framing.
- **Unowned:** the pain appears commercially interpretable, but no funded local company or clear commercial owner was found.

Validation status assignments were author-judged. The detailed company-to-theme validation log is not reproduced here, because it is a commercial annex and may contain productized opportunity mappings. Aggregate rates, status definitions, prompts, sample summaries, and locked taxonomies are disclosed.

3.6 Discovery Index

For market m , let H_m be the set of high-volume themes in that market, V_m be the subset classified as VALIDATED, and C_m be the subset classified as CATEGORY-FORMING. We define:

$$DI_m = \frac{|V_m \cup C_m|}{|H_m|}. \quad (1)$$

The Discovery Index is not a measure of truth, product-market fit, or market size. It measures the share of persona-surfaced high-volume pain themes that correspond to funded or category-forming venture activity in the same market. A higher value suggests more venture-saturated coverage of the persona-surfaced opportunity space. A lower value suggests more partial-gap or unowned space, assuming the validation process is complete.

Table 1: Study samples.

Study	<i>n</i>	Type	Composition summary
India	385	B2C	15 cultural cohorts: Marathi, Gujarati, Tamil, Kannada, UP Hindi belt, Punjabi, Bihar Hindi belt, Bengali, Telugu, Sindhi, Northeast, Malayali, Marwari, Kashmiri, Rajasthani. Ages 18–75. 183 women and 202 men. Metro, tier-2, tier-3, rural, native, migrant, and diaspora-born strata.
UAE	103	B2C	88% expat and 12% Emirati native. Emirati, Filipino, Bangladeshi, Pakistani, Indian regional cohorts, Levantine, Egyptian, Yemeni, Sudanese, Jordanian, and Western expat strata.
Australia	385	B2C	Anglo-Australian dominant sample plus Chinese, Indian, Filipino, British, Greek, Italian, Korean, Pacific Islander, Lebanese, Sudanese, Indigenous, Russian-Australian, and Vietnamese strata. Major metro, regional, and rural-remote coverage.
Southeast Asia	385	B2C	Philippines (140), Malaysia (138), and Thailand (94). Country-internal strata included Tagalog, Cebuano, Ilocano, Bicolano, Waray, OFW-family, Chinese Filipino, Moro Muslim Mindanao, Malay urban/rural, Chinese Malaysian, Indian Malaysian Tamil, Orang Asli, Sarawak/Sabah Bumiputera, Bangkok, Isan, Chinese Thai, Lanna, and Southern Thai.
Germany	175	B2B	Finance and compliance professionals: Finanzleiter, Buchhalter/Controller, Geschäftsführer KMU, Compliance Officers, CFOs, Geschäftsführer, and Steuerberater. 75% German native, 11% EU western migrant, with spread across major and Mittelstand-anchored cities.

4 Results

4.1 Headline results

Across five studies, 1,433 personas surfaced 212 distinct pain themes. High-volume theme counts ranged from 14 to 30 per study. The Discovery Index ranged from 40% to 79%. Table 2 is the centerpiece comparison.

The values do not move randomly. India B2C returned the highest Discovery Index at 79%. UAE, Australia, and Southeast Asia clustered tightly at 40–43%. Germany B2B sat between those values at 58%. This ordering is consistent with a market-saturation interpretation: a high Discovery Index means that more persona-surfaced themes are already visibly funded, while a lower index means that more high-volume themes remain partially owned or unowned.

This interpretation is suggestive rather than conclusive. The study has one sample per geography and validation was author-judged. Still, the pattern is directionally coherent across culturally distinct markets and across one B2B population type.

4.2 Study-level summaries

India B2C. The India panel contained 385 personas across 15 cultural cohorts. It produced 47 distinct themes, 24 of which crossed the high-volume threshold. Nineteen of those 24 were validated or category-forming, giving a Discovery Index of 79%. Three high-volume themes were classified as category-forming because relevant funded activity occurred after the model cutoff. Coverage was 72.5%.

UAE B2C. The UAE panel contained 103 personas, with 88% expat and 12% Emirati native composition. It produced 36 themes, 14 high-volume themes, and a Dis-

covery Index of 43%. Coverage was 94.2%, the highest among the five studies, likely reflecting both smaller sample size and a more demographically concentrated pool.

Australia B2C. The Australia panel contained 385 personas across Anglo-Australian and minority cultural strata, with metro, regional, and rural-remote representation. It produced 50 themes, 21 high-volume themes, and a Discovery Index of 43%. Coverage was 66.0%.

Southeast Asia B2C. The Southeast Asia panel contained 385 personas from the Philippines, Malaysia, and Thailand. It produced 60 themes, 30 high-volume themes, and a Discovery Index of 40%. Coverage was 66.0%. This study produced the strongest cultural self-stratification signal, discussed below.

Germany B2B. The Germany panel contained 175 finance and compliance professionals. It produced 60 or more raw themes compressed into 19 high-volume themes. Eleven of the 19 were validated, giving a Discovery Index of 58%. Coverage was 84.0%. The result suggests that the method generalizes beyond B2C consumer panels into role-specific professional discovery.

4.3 Status distribution

Table 3 reports validation statuses. India has the largest validated/category-forming share. Southeast Asia has the largest unowned count. Germany has a relatively high validated share but also multiple partial gaps in Mittelstand-oriented finance and compliance workflows.

4.4 Cultural self-stratification in Southeast Asia

The Southeast Asia study pooled personas from the Philippines, Malaysia, and Thailand. The prompt did not ask

Table 2: Five-study comparison. Category-forming themes are counted as validated for Discovery Index calculation.

Study	Type	n	Distinct themes	High-volume themes	Discovery Index	Gap/unowned
India	B2C	385	47	24	79%	21%
UAE	B2C	103	36	14	43%	57%
Australia	B2C	385	50	21	43%	57%
Southeast Asia	B2C	385	60	30	40%	60%
Germany	B2B	175	60+	19	58%	42%
Total/range	4 B2C + 1 B2B	1,433	212	14–30	40–79%	21–60%

Table 3: High-volume theme status counts.

Study	Validated	Category-forming	Partial gap	Unowned
India	16	3	4	0
UAE	6	0	8	0
Australia	9	0	10	2
Southeast Asia	12	0	13	5
Germany	11	0	7	1

physical operations, or cross-system glue. To preserve the commercial value of the detailed hypotheses, this paper reports the pattern at the layer level rather than reproducing the full company-to-theme mapping.

Table 4: Country self-stratification in the Southeast Asia mixed panel.

Theme	Dominant country	Share
Remittance / OFW	Philippines	90%
Motorcycle taxi safety	Philippines	92%
Sari-sari operations	Philippines	100%
Prayer logistics	Malaysia	73%
Ramadan fasting logistics	Malaysia	73%
Banana-leaf kitchen preservation	Thailand	73%
Motorbike / scooter culture	Thailand	56%

for country-specific themes, yet high-volume themes self-stratified by country in culturally coherent ways. Table 4 shows the strongest examples.

This matters because a generic regional prompt such as “name Southeast Asian consumer pain points” would be expected to produce broad tropes unless explicitly instructed to stratify by country. In this study, country-specific themes emerged from persona conditioning and open-ended elicitation. We do not report a controlled direct-prompt baseline in this draft; that comparison should be included in a preregistered replication.

4.5 The wrong-layer pattern

The most commercially interesting repeated pattern was not merely that many themes mapped to funded categories. It was that partial-gap themes often sat inside already-funded categories. Personas named a layer adjacent to the legible layer solved by incumbents.

We describe this as the wrong-layer pattern. In market after market, the funded category solved the visible layer: placement, recruitment, discovery, digitization, workflow software, or compliance tooling. Personas named the lived operational layer: capability, coordination, accountability,

Table 5: Abstracted wrong-layer pattern across studies. Specific company mappings and product hypotheses are withheld as a commercial annex.

Market	Funded layer	Persona-named layer	Repeated signal
India	Domestic-service access	Capability inside household work	Not only more supply; capability after access.
UAE	Household-staff recruitment	Ongoing coordination of household labor	Post-placement workflow rather than recruitment alone.
Australia	Local-service discovery	Accountability after selection	Enforcement and reliability rather than provider discovery.
Southeast Asia	Small-retailer digitization	Physical-operational friction	Daily store operations rather than digital inventory alone.
Germany	Accounting/compliance software	Cross-system and authority-interface friction	Connective tissue among tools rather than one replacement tool.

This pattern is compatible with, but not proof of, the claim that culturally grounded persona construction elicits context-specific frictions that direct prompting tends to flatten. It is also compatible with a more conservative interpretation: well-conditioned LLM completions can recombine latent local knowledge into plausible market hypotheses, and external validation is required to separate signal from hallucination.

5 Discussion

5.1 What the Discovery Index measures

The Discovery Index should be read as a market-context measurement, not as a model leaderboard. A high score does not mean that the personas are better in that market. It means that a higher share of high-volume persona-surfaced themes can already be matched to funded or category-forming local activity. A lower score does not mean failure. It may mean the market has more unowned or partially owned commercial whitespace.

This interpretation makes the India result especially informative. India produced the highest Discovery Index because many high-volume consumer pains mapped to funded or category-forming activity. UAE, Australia, and Southeast Asia clustered lower, suggesting less saturated coverage of persona-surfaced pain themes. Germany B2B landed in the middle, consistent with a mature but uneven B2B software landscape.

5.2 Why discovery is distinct from validation

Validation asks whether a known stimulus produces a favorable response. Discovery asks what problem space a population names when no solution is supplied. This difference matters because synthetic users are most vulnerable when they are treated as substitutes for real respondents in decisions that require behavioral evidence. A discovery instrument has a different burden: it must produce candidate hypotheses worth investigating, not final truth.

The method in this paper therefore has a lower but

cleaner claim. The personas surfaced themes. Those themes were classified. High-volume themes were checked against external venture-market evidence. The resulting Discovery Index is measurable. The detailed hypotheses still require real-world research, founder interviews, customer discovery, or market testing.

5.3 Implications for the skeptical literature

The results do not refute skepticism about synthetic users. They narrow the debate. Critics are right that LLM completions are not people and should not be treated as human evidence. However, this study suggests that synthetic personas may still be useful when three conditions hold: the personas are deeply conditioned, the prompt elicits open-ended discovery rather than evaluative approval, and every output is validated against an external reference.

This shifts synthetic personas from “fake respondents” toward “structured hypothesis generators.” That shift is important. It makes the method less grandiose, but more defensible.

5.4 Reproducibility and withheld components

The experiment is reproducible in principle because the prompts, validation status definitions, aggregate results, and locked taxonomies are disclosed. However, three components are withheld or only described at high level: the 16-section seed schema, the voice-discipline mechanism, and the detailed company-to-theme validation log. This is a tradeoff between academic transparency and commercial protection.

The most important next step is independent replication. An ideal replication would use the same prompts, fresh persona samples, the locked taxonomies, multilingual embedding classification, and independent dual-reviewer validation. It would also test a direct-prompt baseline and at least one alternate base model.

6 Limitations

Synthetic personas are not human respondents. The outputs are LLM completions conditioned on persona profiles. They do not establish demand, willingness to pay, behavioral adoption, or causal truth.

Validation was author-judged. Theme validation was conducted by the author using public market evidence. This creates judgment risk. A second-pass independent validation with blinded reviewers is necessary before using exact rates in high-stakes commercial claims.

Single sample per geography. Each geography has one sample. The stability of the Discovery Index across fresh samples is unknown. UAE is especially underpowered at $n = 103$.

The taxonomy was developed iteratively before being locked. The locked taxonomy enables future replication, but the v1.0 taxonomy was produced through analysis of the same runs reported here. Future studies should preregister the taxonomy before execution.

TF-IDF classification is limited. TF-IDF is transparent but less semantically robust than multilingual sentence embeddings. Some unmatched responses may have been semantically close to existing themes.

No controlled direct-prompt baseline is reported. The paper discusses why persona conditioning may matter, but it does not include a locked baseline comparing generic LLM prompts against persona-conditioned outputs. That baseline should be added in the next version.

Venture-market validation is an imperfect external target. Funded startup activity is not equivalent to social importance or market truth. It is a useful external proxy for commercial salience, but it misses bootstrapped businesses, public-sector solutions, informal-sector workarounds, and local non-venture categories.

7 Conclusion

We reported an inverse-docking experiment for culturally calibrated synthetic personas. Across five studies, 1,433 personas produced 212 distinct pain themes. Between 40% and 79% of high-volume themes mapped to funded or category-forming venture activity in the relevant market; the remaining 21% to 60% were partial-gap or un-owned. The Discovery Index varied coherently across markets, and the method generalized from B2C consumer panels to a Germany B2B finance and compliance panel.

The strongest finding is the repeated wrong-layer pattern. Personas did not simply reproduce broad startup categories. They often surfaced adjacent frictions inside already-funded categories, where incumbents addressed a visible layer but not the lived operational layer. This is where the discovery use case differs from the validation use case.

The claim is deliberately bounded. Synthetic personas do not replace real research. They can, however, produce measurable and externally checkable discovery hypotheses

when culturally grounded, prompted without a stimulus, classified into locked taxonomies, and validated against external market evidence. We position this work as the inverse complement to protocol-docking studies of synthetic personas: instead of measuring fidelity to known human data, it measures open-ended discovery against a changing external market.

Data Availability

The 1,433 persona responses, locked taxonomies, classifier code, and validation logs are available on request to qualified researchers under appropriate confidentiality conditions. The full persona seed-generation method and voice-discipline mechanism are not included in this paper and are planned for a separate technical report.

Ethics Statement

No human-subject survey or interview data was collected for this study. All respondent outputs were generated by synthetic personas. External validation relied on public information about companies and markets. Because the author is also the founder and CEO of TwinSim, readers should treat commercial claims with appropriate caution until independently replicated.

Conflict of Interest

The author is the founder and CEO of TygrX Inc. / TwinSim and has a direct commercial interest in the system evaluated in this paper.

A Prompts

A.1 B2C prompt

Given your life, which problem have you faced where the problem is not just obvious, it is so common that it is not even considered a problem to be solved? Avoid extremely common problems like better air conditioner. Things like cracked heels or women in India are true unique problems, take that as example and answer.

A.2 B2B prompt

Given your professional role, name a problem so common in your daily work that nobody considers it solvable. Avoid generic complaints about software or meetings. Think about the friction you have stopped noticing because it has always been there.

B Locked Taxonomies: High-Volume Themes

The tables below reproduce theme names, counts, and statuses. They intentionally omit regex patterns, anchor descriptions, company mappings, and gap rationales.

B.1 India B2C

Table 6: India B2C high-volume taxonomy.

ID	Theme	<i>n</i>	Status
IN01	Domestic help reliability	40	Category-forming
IN02	Regional food and ingredients	31	Validated
IN03	Traffic, commute, transit	25	Validated
IN04	Parking, urban	24	Validated
IN05	Water supply shortage	22	Validated
IN06	Mental health and stress	22	Validated
IN07	Standing-worker footwear	18	Partial gap
IN08	Government document concierge	16	Validated
IN09	Jain, halal dietary filter	16	Validated
IN10	Public toilets, washrooms	15	Validated
IN11	Senior small-print readability	13	Partial gap
IN12	Regional language content	12	Category-forming
IN13	SME finance, GST	12	Validated
IN14	Medication packaging, elder	9	Category-forming
IN15	Grocery, provisions	9	Validated
IN16	Digital payment, banking	8	Validated
IN17	Rural emergency patient transport	8	Partial gap
IN18	Childcare, school admin	7	Validated
IN19	Skin, hair, personal care	7	Validated
IN20	Elder care support	6	Validated
IN21	Garbage, waste management	5	Validated
IN22	Education, career, skill bridge	5	Validated
IN23	Gig worker phone battery	5	Partial gap
IN24	Fisherman cold storage at sea	5	Unowned

B.2 UAE B2C

Table 7: UAE B2C high-volume taxonomy.

ID	Theme	n	Status
AE01	Remittance friction	30	Validated
AE02	Authentic regional food	14	Validated
AE03	Healthcare worker pain	9	Partial gap
AE04	Gig delivery worker friction	8	Validated
AE05	Majlis, social obligation	7	Partial gap
AE06	Household staff coordination	7	Partial gap
AE07	Prayer schedule logistics	6	Partial gap
AE08	Blue-collar labor camp life	5	Partial gap
AE09	Emirati home-cooked food subscription	5	Partial gap
AE10	Government paperwork attestation	5	Validated
AE11	Home repair, maintenance trust	5	Validated
AE12	Public bus, blue-collar commute	4	Partial gap
AE13	Arabic-first software gap	4	Partial gap
AE14	Visa documentation processing	4	Validated

B.3 Australia B2C

Table 8: Australia B2C high-volume taxonomy.

ID	Theme	n	Status
AU01	Tradie reliability and accountability	47	Partial gap
AU02	Public transport reliability	35	Validated
AU03	Kids activity organization	23	Validated
AU04	Rural and remote isolation services	22	Partial gap
AU05	Traffic, commute, parking	22	Validated
AU06	Mental health and stress	21	Validated
AU07	Kids school paperwork chaos	18	Partial gap
AU08	Kid school dropoff carpool	18	Partial gap
AU09	Cost of living, grocery	16	Validated
AU10	Garden tools seniors ergonomic	15	Partial gap
AU11	Dust and red dirt regional	11	Partial gap
AU12	Family kids schedule (blended)	10	Partial gap
AU13	Supermarket self-checkout friction	10	Unowned
AU14	Regional internet, mobile reception	10	Validated
AU15	Car wash, dust, Perth/outback	8	Unowned
AU16	Share house internet, chores	6	Partial gap
AU17	Late-night student healthy food	6	Validated
AU18	Meal prep, busy professional	6	Validated
AU19	Outdoor lifestyle friction	6	Partial gap
AU20	Diaspora authentic food groceries	5	Validated
AU21	Australian-specific pests	5	Partial gap

B.4 Southeast Asia B2C

Table 9: Southeast Asia B2C high-volume taxonomy.

ID	Theme	n	Status
AP01	Public transport, traffic	30	Validated
AP02	Remittance, OFW (PH dominant)	29	Validated
AP03	Motorbike, scooter culture (TH dominant)	27	Validated
AP04	Motorcycle taxi safety (PH dominant)	24	Validated

ID	Theme	<i>n</i>	Status
AP05	Government paperwork ASEAN	18	Partial gap
AP06	Humidity, mould, housing	16	Partial gap
AP07	Sari-sari coin management (PH only)	15	Partial gap
AP08	Prayer logistics workplace (MY dominant)	15	Validated
AP09	School kids admin and costs	15	Validated
AP10	OFW family separation (PH only)	14	Partial gap
AP11	Cultural food authentic	14	Validated
AP12	Reading glasses seniors	13	Partial gap
AP13	Banana leaf kitchen preservation (TH dominant)	11	Partial gap
AP14	Ramadan fasting logistics (MY dominant)	11	Validated
AP15	Gig delivery (Grab, foodpanda)	9	Validated
AP16	Indigenous Orang Asli, Lumad	9	Unowned
AP17	Healthcare worker pain	8	Partial gap
AP18	Balikbayan, diaspora packages	8	Validated
AP19	Wet factory shoes rainy season	6	Unowned
AP20	Wet market fish freshness	6	Unowned
AP21	Elder care, aging parent ASEAN	6	Partial gap
AP22	Public toilets, traffic-stuck commuters	6	Unowned
AP23	Chinese minority SE Asia	6	Partial gap
AP24	Flooding, rain, typhoon	5	Partial gap
AP25	Motorcycle rider climate protection	5	Partial gap
AP26	Traffic parking APAC	5	Validated
AP27	Digital payment multi-QR	5	Validated
AP28	Regional dialect translation	5	Partial gap
AP29	Gig worker phone battery	5	Partial gap
AP30	Traditional kuih cooking pain	5	Unowned

B.5 Germany B2B

Table 10: Germany B2B high-volume taxonomy.

ID	Theme	<i>n</i>	Status
DE01	DATEV/ERP reconciliation	30	Validated
DE02	VAT and USt compliance	26	Validated
DE03	Compliance documentation, audit	20	Validated
DE04	Excel manual consolidation	15	Partial gap
DE05	Mittelstand software size gap	14	Partial gap
DE06	GDPR DSGVO workflow	13	Validated
DE07	Legacy ERP integration glue	13	Partial gap
DE08	Handwerk SMB paper-to-digital	11	Partial gap
DE09	Paper-to-digital friction	10	Validated
DE10	Steuerberater client document collaboration	9	Partial gap
DE11	Finanzamt and authority communications	8	Partial gap
DE12	Digital signature friction	8	Validated
DE13	Tax compliance workflow	7	Validated
DE14	Supplier invoice format chaos	7	Validated
DE15	Cross-border EU tax	7	Validated
DE16	Bundesland portal fragmentation	6	Unowned
DE17	Master data quality	6	Validated
DE18	Three-way match PO-invoice	5	Validated
DE19	Tool sprawl, too many apps	5	Partial gap

C Coverage and Binomial Intervals

The intervals in Table 11 are simple Wilson intervals over high-volume theme counts. They should not be treated as definitive inferential intervals because themes are not independent human respondents and the validation process is author-judged. They are included only to show the uncertainty implied by small denominators.

Table 11: Discovery Index counts and Wilson intervals.

Study	k	n	DI	Wilson 95% interval
India	19	24	79.2%	59.5–90.8%
UAE	6	14	42.9%	21.4–67.4%
Australia	9	21	42.9%	24.5–63.5%
Southeast Asia	12	30	40.0%	24.6–57.7%
Germany	11	19	57.9%	36.3–76.9%

D Validation Procedure

For each high-volume theme, validation proceeded as follows: (1) construct English and local-language search queries for the theme; (2) search public funding news, startup databases, and company directories; (3) verify company activity on direct company websites where possible; (4) classify the theme as validated, category-forming, partial gap, or unowned; (5) record the rationale in the validation log. The company-to-theme validation log is not reproduced in this paper.

References

- [1] Joon Sung Park, Joseph C. O’Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative Agents: Interactive Simulacra of Human Behavior. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. arXiv:2304.03442.
- [2] Lisa P. Argyle, Ethan C. Busby, Nancy Fulda, Joshua Gubler, Christopher Rytting, and David Wingate. 2023. Out of One, Many: Using Language Models to Simulate Human Samples. *Political Analysis* 31(3):337–351. arXiv:2209.06899.
- [3] Gati V. Aher, Rosa I. Arriaga, and Adam Tauman Kalai. 2023. Using Large Language Models to Simulate Multiple Humans and Replicate Human Subject Studies. In *Proceedings of the 40th International Conference on Machine Learning*. arXiv:2208.10264.
- [4] Jorn K. Teutloff. 2025. Synthetic Founders: AI-Generated Social Simulations for Startup Validation Research in Computational Social Science. arXiv:2509.02605.
- [5] DeepSeek-AI. 2024. DeepSeek-V3 Technical Report. arXiv:2412.19437.
- [6] Nielsen Norman Group. 2024. Synthetic Users: If, When, and How to Use AI-Generated Research. NN/g, June 21, 2024.
- [7] W. Michelle Harris. 2025. The Synthetic Persona Fallacy: How AI-Generated Research Undermines UX Research. *ACM Interactions*, December 17, 2025.
- [8] MeasuringU. 2026. A Review of Experiments with Synthetic Users.
- [9] E. Kuric. 2026. Synthetic Participants Generated by Large Language Models: A Systematic Literature Review of 182 Studies. Research Square preprint.