Journal of Quantitative Description: Digital Media 5(2025), 1–57 10.51685/jqd.2025.014

Information on Integration: Administrative Offerings & Immigrant Engagement —An Explorative Large-N Analysis of the 'Integreat' Platform in Germany

SAMIR KHALIL[1]
German Center for Integration & Migration Research, Germany
& University of Potsdam, Germany

Which topics do local administrations focus on when providing refugees and other immigrants with information about their first steps after arrival – and how well are they received? These questions are particularly important in federal systems, where regions bear significant responsibility for integration efforts, despite limited resources and a complex network of offices to navigate. This study explores "Integreat", Germany's largest multilingual information platform designed to facilitate integration at the district and city levels. Using the platform's full website visit history from 2018–2024, it analyzes several million records from 100+ participating regions. Results document significant differences in the thematic focus across districts. While some topics, such as work and training, are widely available, others are much less frequently covered. Multivariate analyses comparing topic availability and the likelihood of visiting related pages reveal several key insights. Pages offering immediate practical value (opening hours; mobility) are highly visited when available, yet absent in many districts. Topics related to acquiring essential skills or knowledge (e.g., language learning) show intense engagement relative to supply and users readily navigate detailed pages to access them. Finally, engagement differs across translations (e.g., Arabic, Ukrainian) for given topic availability, reflecting the diverse needs of different immigrant groups. The work concludes by deriving implications for administrative practice and noting the potential of these data for further descriptive and inferential research on refugee and immigrant integration.

Keywords: integration, multilingual, local government, immigration, refugees

[1]Samir Khalil: khalil@dezim-institut.de
Date submitted: 2025-02-12

Introduction

Many countries face a structural dilemma concerning optimal information provision to refugees and immigrants: On the one hand, central governments often struggle to provide relevant information because integration processes largely occur at the local level, where structures vary significantly between regions and change frequently over time, particularly in decentralized countries like the United States or Germany (Tjaden & Spörlein, 2023). On the other hand, local administrations themselves often lack the financial and technical resources needed to deliver adequate multilingual information to refugees and immigrants in their jurisdictions. In particular, they frequently do not have the necessary IT infrastructure to effectively reach their target audiences digitally.

This is critical because the internet and social media serve as a primary channel through which refugees and immigrants gather information (Stapf, 2019). During their journeys, online research helps refugees avoid dangers and navigate more safely, and many intensify their internet usage significantly throughout their migration process (Borkert et al., 2018; Jauhiainen et al., 2021; Merisalo & Jauhiainen, 2019). Upon arrival, digital tools and services play a crucial role in supporting integration efforts, such as through online video platforms, digital language courses, and apps tailored to newcomers (Bradley et al., 2017; Jurić, 2022; Kukulska-Hulme et al., 2015). Previous studies highlight refugees' intense engagement with information that originates from official sources, emphasizing the need for local administrations to meet this demand (Siegel et al., 2024).

The present research project examines "Integreat," a digital information platform in Germany that equips city and district administrations with a unified digital infrastructure (Tür an Tür - Digitalfabrik gGmbH, 2023). This platform allows authorities to independently provide refugees and immigrants with locally relevant information, accessible via mobile apps or web browsers, and available in multiple languages (Figure 1). Administrations primarily focus on refugees, with 82% offering Arabic translations and 76% offering Ukrainian translations. By May 2024, a total of 644 region-language combinations were available online, distributed across 105 regional apps. In the following, these combinations will be referred to as *platform versions*.

The research questions stem from a key limitation faced by participating regions: the need to independently create, maintain, and update their platform versions with content, which requires significant personnel resources (Tür an Tür - Digitalfabrik gGmbH, 2024). Consequently, this study explores (1) *which topics are prioritized given these limited resources*; (2) *how this prioritization aligns with users' information demands*; and (3) *whether the supply-demand relationship varies by language*.

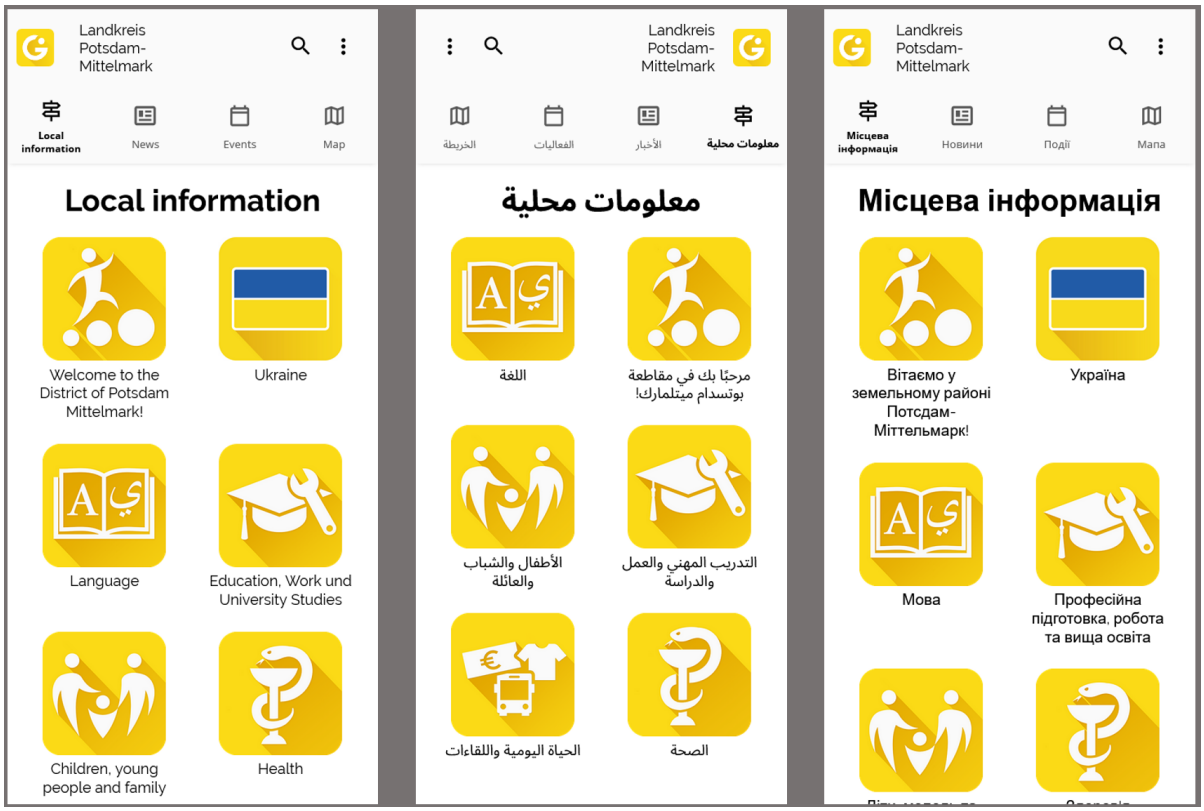

Figure 1. Landing pages for district Potsdam-Mittelmark in English, Arabic, and Ukrainian: Landing pages of three exemplary platform versions (region-language combinations) out of 644 observed in this study.

To address these questions, the study examines the platform's complete page visit history. By processing over 11 million records from 2018 to 2024 across more than 100 participating regions, it offers a detailed view of how regions prioritize integration topics and translations. In addition, the study employs multivariate logistic regression models to assess how variations in content supply across regions are associated with users' engagement intensity.

The study finds major differences between participating regions, both in topic coverage and topic depth. If topics are covered, an increase in topic quantity is positively associated with user engagement, but the strength of this relationship varies across topics. In addition, the demand for topics varies by access language, likely reflecting the distinct needs of migrant groups.

This study offers detailed descriptive insights into an example of how local administrations allocate scarce resources to integration work and how these decisions align with the needs of refugees and immigrants. It makes use of novel digital trace data revealing the responsiveness to official com-

munication on integration at the local level.[2] These findings support the design of more targeted and effective communication strategies and provide a foundation for future research on the impact of digital information access on integration outcomes.

Context

Initial lack of language proficiency

The vast majority of refugees and even most highly skilled immigrants arrive with low knowledge of the German language (Kosyakova et al., 2021; Liebau & Romiti, 2014). In their year of arrival, only 15% of refugees report good or very good German reading skills.[3] Therefore, mechanisms that positively affect language acquisition – such as language courses, contact with locals, residential contexts, and the duration of stay in reception centers – unfold their impact only some time after arrival (Kanas et al., 2022; Khalil & Tjaden, 2025; Khalil et al., 2022; Marbach et al., 2025).

Initially, most refugees and immigrants must instead overcome significant language barriers to manage their first steps, such as registering with the residents' registration office, finding a language course, applying for a work permit, or obtaining recognition of qualifications acquired abroad. According to the German IT Planning Council, foreign nationals rarely use online resources for administrative services–even if available–due to incomplete translations or fragmented supply across multiple sites (IT Planning Council, 2017, p. 15).

Regional peculiarities hinder effective communication

As in many other countries, refugees in Germany are assigned to regions through distribution systems that do *not* consider individual characteristics. This lack of choice may exacerbate initial disorientation, as refugees are unable to move closer to friends or relatives who could provide essential support (Battisti et al., 2022; Bundesamt für Migration und Flüchtlinge, 2022). In addition, the regions are characterized by very different structural conditions, for example, in terms of economic strength, urbanity or social infrastructure (Federal Ministry for Economic Affairs and Climate Action, 2024; Federal Ministry of the Interior and Community, 2019; Fink et al., 2019; Khalil & Tjaden, 2025; Khalil et al., 2022; Ostermayer

[2]The underlying data are not publicly available and may be requested from Tür an Tür - Digitalfabrik gGmbH (2023).

[3]See Figure A1 in the supplemental materials for a descriptive analysis, based on the IAB-BAMF-SOEP study on refugees. Until the fifth year after arrival, this share rises to 63%.

et al., 2022). However, what is particularly obstructive to efficient communication is the decentralized nature of most organizations involved in integration support and immigrant administration.

First, many essential integration measures are implemented by local actors–even when policies are formally regulated at higher national or state levels. For instance, integration and language courses–which are the most cost-intensive political measures to promote integration, costing up to approximately one billion euros annually–are implemented by a diverse range of local organizations.[4] As a result, the availability of specific contact persons and the quality of online course information can vary significantly depending on the location. Similarly, immigration offices (*Ausländerbehörden*), which serve as the primary contact points for refugees and immigrants on their residence status, are typically integrated into the local administrative structures of districts or cities (Eule, 2017). Consequently, appointment scheduling is location-specific and may require non-standardized forms, emailing designated staff, or in-person visits.

Second, the refugee migration of 2015 and 2016 placed significant pressure on local officials and civil society when around one million refugees, mainly from Syria, Afghanistan, and Iraq, settled in Germany. The situation led to numerous initiatives: Many cities and districts have developed integration concepts (*Integrationskonzepte*) with self-imposed goals[5] and are enhancing coordination by appointing local integration officers (*Integrationsbeauftragte*) or creating dedicated integration departments (Filsinger & Gesemann, 2017). Kuhlmann et al. (2024) argue that the influx exerted external pressure on administrative systems, fostering greater vertical interdependence across different administrative levels and prompting municipalities to supplement nationwide integration efforts.

Third, local civic engagement in refugee support experienced a remarkable increase following the arrival of refugees. Lueders (2023) found that the number of new refugee support associations surged from previously 15 per year to 125 in 2015 (Lueders, 2023, p. 9).[6] These associations often undertake central tasks, including the initial care of refugees upon arrival, providing training measures, conducting counseling sessions, and creating various opportunities for refugees to interact with local residents. Moreover, empirical studies have demonstrated a robust correlation between local civic engagement and refugees' contact with Germans, language proficiency, and life satisfaction (Barreto et al., 2022;

[4]These include, for instance, adult education centers (*Volkshochschulen*), specialized language schools, or church organizations (Aumüller, 2017; BAMF, 2024; Mediendienst Integration, 2025).

[5]These self-imposed goals may include comprehensive health and psychological support or limiting time in collective housing (Filsinger & Gesemann, 2017; Khalil & Tjaden, 2025).

[6]This surge led to a doubling of the absolute number of such organizations, reaching approximately 2.5 organizations per 1,000 inhabitants.

Priemer & Schmidt, 2019).

Overall, initiatives by community organizations and local officials highlight the essential capacity to address migration-related challenges. Nevertheless, when combined with traditionally decentralized state structures, the result is an exceptionally heterogeneous and dynamic integration infrastructure that is difficult to keep track of, especially for newcomers who are not familiar with the local language and procedures.

Multilingual information and integration

Conceptually, refugees' and immigrants' access to locally relevant and multilingual information can be considered part of their *navigational integration*, meaning the *"ability to manage basic needs in the host country, such as seeing a doctor, addressing legal problems, and searching for jobs"* (Harder et al., 2018, p. 11484). Additionally, providing information in familiar languages may ease anxiety and uncertainty and enhance trust in new systems (de Alpuim-Gonçalves et al., 2025; Potter et al., 2024).

Empirical evidence on how official information provision is associated with immigrants' engagement or integration remains limited. Siegel et al. (2024) analyze several thousand Facebook posts from Arabic-language pages targeting Syrian refugees and find that posts produced by or linking to official sources consistently garnered more engagement than those from unofficial sources. Tjaden et al. (2022) demonstrate that Arabic, Turkish, and Russian speakers living in Germany were 1.2 to 2.4 times more likely to engage with advertisements promoting COVID-19 vaccinations when the information was presented in their native languages, compared to German. Similarly, an Australian RCT showed that a multilingual, culturally adapted website increased depression knowledge and reduced personal stigma among Greek- and Italian-born immigrants (Kiropoulos et al., 2011).

Case and research question

Concept

Integreat is an open-source platform that offers tailored local information and services for refugees and immigrants in Germany. Its content is curated by local authorities and is available in the languages most relevant to each community. The platform was developed by *Tür an Tür - Digitalfabrik gGmbH*, a non-profit enterprise (Tür an Tür - Digitalfabrik gGmbH, 2023). With more than 100 district and city administrations participating by 2025, it represents the most widespread multilingual platform with local

information for immigrants in Germany, presumably even across Europe.[7]

Users access content on the platform using mobile apps (Android, iOS) or directly via web browsers. Initially, they need to select the region and language relevant to them. The platform covers topics such as social services and administrative procedures (e.g., language courses, health services, childcare). Importantly, local authorities are independently responsible for the content of their platform versions, meaning that they must allocate internal resources to provide content and translations. As a result, the availability can vary significantly across regions, which may reflect distinct integration environments but also regions' priorities or available resources.

Limited resources of participating regions

Setting up and maintaining the application initially entails a considerable amount of extra work for city and district administrators. This effort is often taken on by employees who originally performed other tasks, for instance, local integration managers, migrant advisory boards, or secretaries. According to the annual user survey from 2024, 58.3% reported investing less than four working hours per week to update the app, and only 10% reported investing more than eight hours per week (Tür an Tür - Digitalfabrik gGmbH, 2024). It was reported in comments that the time available to maintain the application depends heavily on the scope of other tasks and that the time that would be necessary for a satisfactory provision of information often cannot be found (ibid.). The task of providing information thus represents one of several areas of responsibility in the context of local integration that potentially suffer from a lack of resources.[8]

Research questions

Building on the challenges posed by the limited resources of participating regions, this study explores three interrelated questions in the following:

[7]Alternative integration solutions in Europe-such as BLEND IN, a mobile orientation toolkit offering language, cultural, and social guidance in Italy, Greece, Malta, and Cyprus, and RefAid, a map-based app connecting refugees with local service providers in Malta-are typically funded as short-term projects that often become inactive once grant funding ceases. By contrast, Integreat is maintained in a decentralized manner by local municipalities that continuously update region-specific information, offering a more sustainable and resilient model for refugee integration.

[8]Other areas that are often discussed include, for instance, the provision of sufficient housing. In 2024, 71.2% of municipalities reported a "challenging but still feasible" situation when accommodating refugees, and 22.9% even reported an "overload in emergency mode" (Kühn & Ziegler, 2024).

(RQ1) Topic supply and prioritization: *Which topics do administrative staff prioritize in their communications about integration, given their typically limited resources available?* This question aims to document regional variations in topic availability and in the proportion that topics represent of their respective local platform version, thereby highlighting both commonalities and differences across administrations.

(RQ2) Alignment with user engagement: *To what extent does the (prioritized) selection of topics align with actual user engagement?* This inquiry examines to what extent the proportional allocation of topics corresponds with the intensity of page views, thereby revealing potential over- or under-representation of certain topics.

(RQ3) Language-specific variations: *How does the relationship between information supply and demand differ across languages?* Immigrant groups face different information needs–for instance, due to varying legal situations or durations of stay. This descriptive question explores how topic engagement differs across provided translations, thereby reflecting these group-specific needs. It also highlights that providing targeted information is challenging for administrations, as it requires both accurate translations and careful assessment of the topics most relevant to each immigrant group. Without sufficient staffing, administrations may rely on simple translations of the German reference platform version, a practice that might be associated with lower relevance and engagement.

Data

To answer the formulated research questions, this study utilizes anonymous website logs from the platform, capturing all individual page visits dating back from its initial implementation in May 2018 until data provision in May 2024. The raw dataset comprises 11.8 million records and three variables: (1) the *page URL*, (2) the exact *date and time of page visits*, and (3) the *user-agent string*, which is sent by web browsers and contains information on the viewers' browser provider, operating system, and device used. As such, the data represent digital trace data that were not originally collected for scientific usage but potentially hold very valuable information for social science research (Cesare et al., 2018; Tjaden, 2021). The data were provided upon request by *Tür an Tür Digitalfabrik gGmbH*.

Sub-Sampling and cleaning

Several reduction steps were implemented to enhance accuracy and relevance as far as possible. Table A3 in the supplemental materials provides an overview of all reduction steps. For the multivariate analyses presented below, 2.96 million cases remain, representing a 74.9% reduction, mainly due to the exclusion of the German platform versions and of data from periods before 08/2021, when information on viewed topics was not yet available.

Descriptive context

Evolution of the platform

This section provides an overview of the platform's geographical reach and its target groups, thereby establishing the necessary context for later analyses. By May 2024, 105 regional apps over 108 districts[9] and cities have been launched, covering 27% of all German regions on the level of districts and cities (Figure 2). Several East German states remain underrepresented, while early adopters are somewhat clustered in West German states Bavaria and North Rhine-Westphalia.[10] User survey responses from administrations often indicate that they first learned about the platform through local networks or neighboring districts (Tür an Tür - Digitalfabrik gGmbH, 2024). Nevertheless, the shortest "walk" across borders between new and already participating regions does not differ significantly from simulated random diffusion (Figure A11; Table A5), which contrasts with these statements and points to supra-regional diffusion of knowledge and platform adoption.

There is a steady increase of about one additional active region per month. Shortly after the start of the invasion of Ukraine, ten regions launched the app until April 2022, likely in response to the information needs of arriving refugees. Overall, platform usage increased substantially from 12.3 thousand page visits in June 2018 to 344.4 thousand visits in April 2024. This growth reflects both more participating regions and higher activity per region, with median visits rising from 884 in 2018-Q4 to 5,571 in 2024-Q1.

The page visits notably surged after the onset of the war in Ukraine. Thus, from 2021-Q4 to

[9]This discrepancy arises from three regional apps that cover two districts each.

[10]East German states: Thuringia (N = 1, 4.6%), Saxony (N = 2, 15.4%), Saxony-Anhalt (N = 2, 14.3%), and Mecklenburg-Vorpommern (N = 0). Of the first 27 districts launching until early 2018, 13 are located in Bavaria and six in North Rhine-Westphalia. Until May 2024, Hesse became the most strongly represented state, with 42% of participating districts.

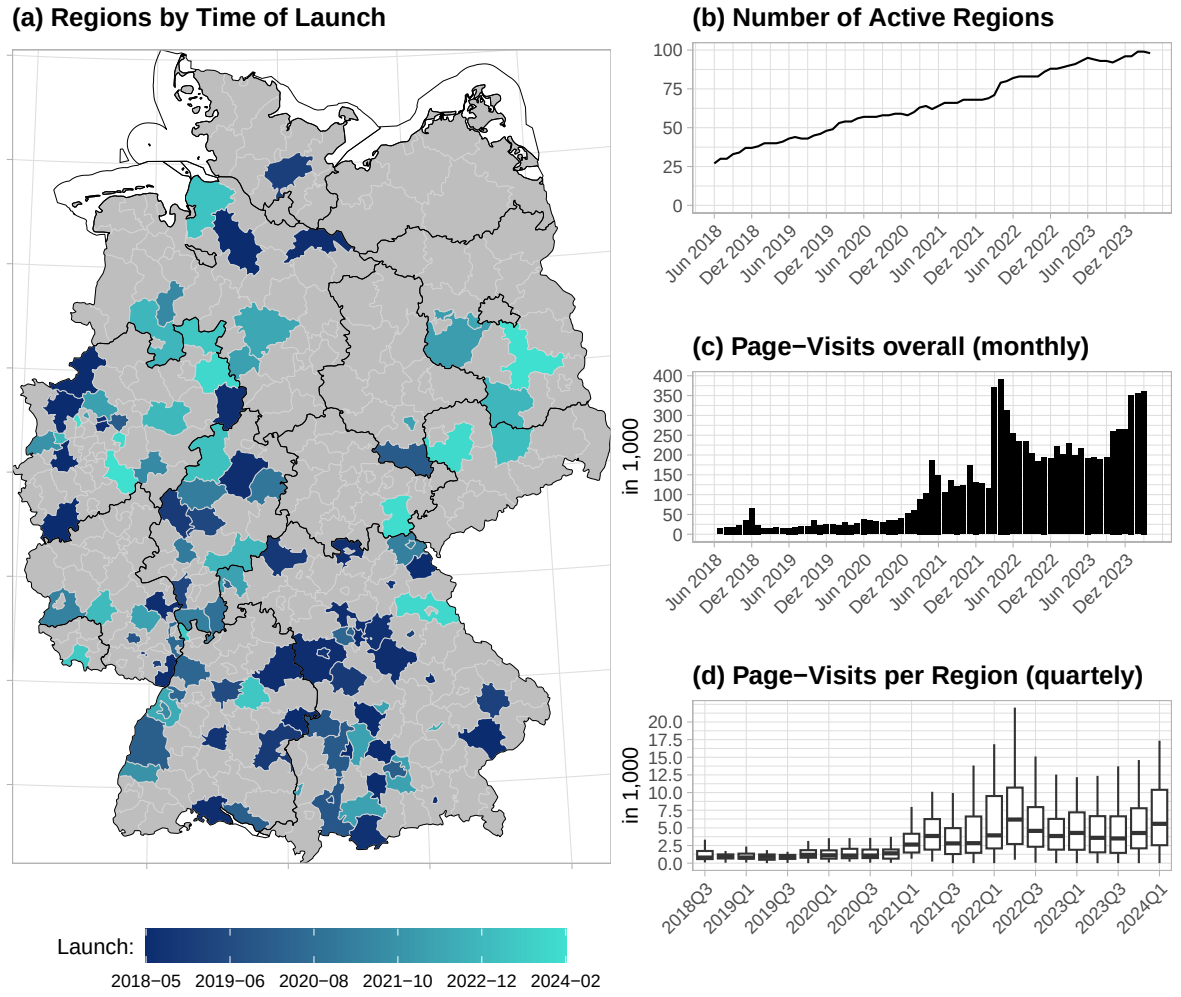

Figure 2. Integration platform – spatial and temporal activity: (a) Time of platform launch across German counties and independent cities (N = 400). Grey areas indicate no launch by 2024. (b) Number of simultaneously active regions (N = 27–99 regions, depending on the month). (c) Total page visits per month (d) Quarterly page visits, summarized as box plots (min, Q25, median, Q75, max), calculated from region-quarter totals (N = 1,480 region-quarters), separately by quarter.

2022-Q2, median page visits per region increased from 2,826 to 6,178 (+118.6%), and the overall page visits increased from 115,000 to 371,000 only between February and March 2022 (+222.6%).

Participating (N = 105) and non-participating (N = 292) regions exhibit largely similar distributions across key regional characteristics, including the share of residents from asylum countries (Figure A2). Urban areas–with higher population density, rents, and incomes–tend to participate slightly

more frequently, though the differences are modest. Similarly, the election results for the far-right and Green parties are distributed in a comparable way, suggesting that the political landscapes in both groups are similar.

Translations

Local administrations independently decide which groups of refugees and immigrants to target by selecting translations that reflect their target groups' native languages. The way they make this decision may be guided by strategic considerations or perceived needs. For instance, they may face limited resources and therefore prioritize languages spoken by the largest immigrant groups or those who are perceived the neediest for information in their mother language. Figure 3 reports the share of the 15 most frequently offered languages from June 2018 to May 2024.

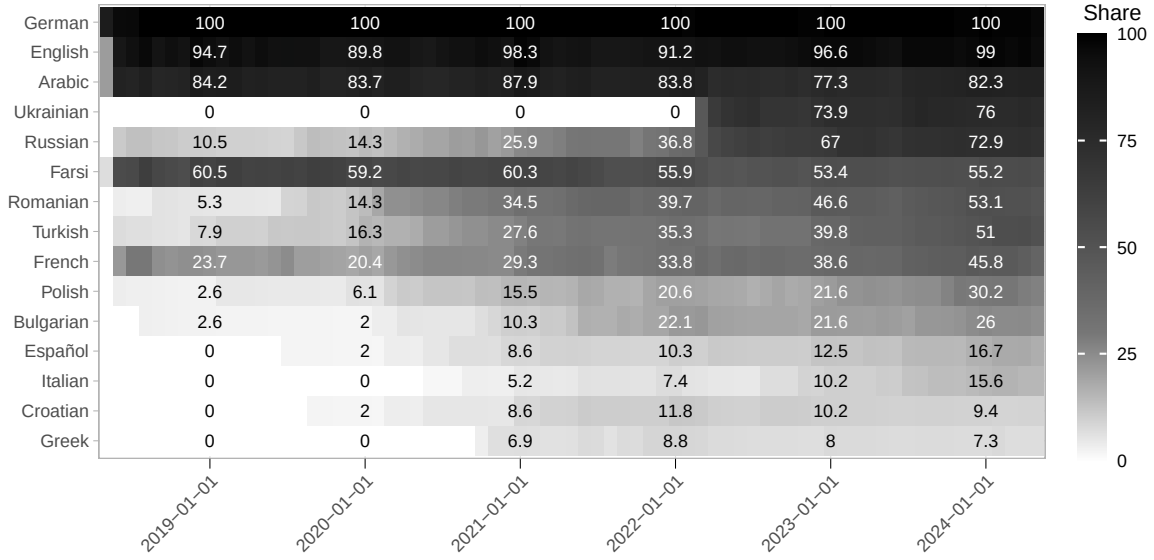

Figure 3. Languages provided (2018-2024): Percentage of regions offering each language over time (top 15 languages as of January 2024). Languages are sorted in descending order based on their share in January 2024.

The German language usually serves as a reference for all others and is therefore always available. Likewise, English translations are practically always available (99%). Further very frequently available languages as of January 2024 are Arabic (82.3%), Ukrainian (76%), and Russian (72.9%), reflecting the languages spoken by the two largest refugee groups in Germany from Syria and Ukraine.

Many languages experienced increasing coverage over the years. Thus, most administrations

reacted quickly to the altered demand after the arrival of refugees from Ukraine. Just three months after the full invasion began, 70% of the regional apps provided Ukrainian and 60% Russian as language options to choose from. Besides, steady increases are observable for Romanian and Bulgarian, which indicate that many regions use the platform increasingly not only to address refugees but also immigrant EU workers.

Topic modelling & empirical strategy

Supply side

To map the supply side, statistics were calculated at the level of distinct URLs (N = 128,420), each annotated with additional information: (1) the *region*, (2) the *display language*, and (3) the viewed *topics*. Topic components of the URL were originally stored as non-standardized, language-specific strings[11] and were harmonized into English using the Google Translate API in R for further classification (El-Desouky, 2023). Subsequently, a systematic keyword search using regular expressions was applied to identify 20 recurring topics,[12] which were then coded as dummy variables. Furthermore, it is assumed that pages were online and accessible during the period between the *earliest* and *latest* observed page visit per URL. Consequently, the *number of available URLs per topic* and the *number of available URLs overall* could be calculated on a daily basis–and separately for all region-language platform versions. This procedure enables the mapping of a detailed longitudinal picture of administrations' emphasis on topics. Finally, the supply-side data were re-merged with individual page visit data based on region, language, and date (daily), providing information on the content available in the same platform version at that time.

Topic selection

The selection of the 20 topics proceeded in a semi-supervised manner. First, candidate topics were derived inductively from the most frequent terms and co-occurrences within URLs (Figure A7). Second, supervised adjustments were implemented by adding topics with sufficient but less prevalent issue areas

[11] E.g., English: "https://integreat.app/potsdam/en/language"; German "https://integreat.app/potsdam/de/deutsch-lernen"; Arabic "https://integreat.app/potsdam/ar/اللغة".

[12] (1) Work & Training, (2) Counseling & Help, (3) Family & Children, (4) Learning German, (5) Childcare & School, (6) Health, (7) Welcome / First Steps, (8) Legal Issues, (9) Info on Region, (10) Housing, (11) Offices, (12) Mobility, (13) Women, (14) Clubs & Sports, (15) LGBTQ, (16) COVID-19, (17) Open. hours & Holidays, (18) Ukraine, (19) Elderly, (20) Racism & Discrimination.

whose importance is well reported in academic literature or governmental reports on integration, such as housing or discrimination (Commissioner for Migration, Refugees and Integration (IB), 2024; Khalil & Tjaden, 2025). The premise here is that less frequently offered topics may signal scarcity, especially when user engagement proves to be disproportionately higher than (low) supply-levels would suggest.

To validate and contextualize this manner of topic modeling, a fully unsupervised topic model (also with N=20) was estimated using the Latent Dirichlet Allocation (LDA) algorithm with Gibbs sampling, as implemented in the *topicmodels* R-package (Grün & Hornik, 2024), specifically designed to work well with very short segments of text such as those present here (Phan et al., 2008). In the supplemental materials, the top-10 most frequent and distinctive terms per LDA-topic are presented (Figure A8; Figure A9) and compared against the chosen semi-supervised RegEx-based model (Figure A10). Strikingly, about three-quarters of the RegEx topics show concentrated signals within a single LDA topic, supporting their distinctiveness and relevance. The remaining quarter shows signals across two LDA topics, suggesting they combine subtopics in which specific terms frequently co-occur. For example, the RegEx category "Racism & Discrimination" shows strong signals in the LDA topic on women, but also moderately elevated signals in the LDA topic on the recognition of professional qualifications.

Further materials on topic modelling are documented in the supplemental materials: (1) Table A1 documents the exact RegEx queries used to identify the topics; (2) Table A2 reports the topics alongside the most frequent words with which they co-occur, providing an evaluation of the identification quality; and (3) Figure A3 presents a coincidence matrix assessing the extent to which topics can be clearly distinguished from one another.

Empirical strategy

The methodological approach differs for the three research questions posed: RQ1 (*topic prioritization by regions*) is analyzed using uni- and bivariate methods at the level of platform versions (i.e., region-language combinations), whereas RQ2 (*alignment with user engagement*) and RQ3 (*language-specific variations*) are examined using multivariate models at the level of individual page visits. Thus, for analyzing RQ2 and RQ3, logistic regression models are performed separately for each of the identified topics (N=20):

$$
\begin{aligned}
Logit(Pr(TopicView_i = 1)) = \; & \beta_0 + \beta_1 TopicSupply_{r,l,t} + \beta_2 Lang_l + \beta_3 (TopicSupply_{r,l,t} \times \\
& Lang_l) + \beta_4 Month_i + \beta_5 Weekend_i + \beta_6 Device_i + \\
& \beta_7 SocialMedia_i + \beta_8 AppSize_{r,l,t} + \beta_9 MonthsLive_{r,l,t} + \\
& \beta_{10} District_r + \epsilon_{i,r,l,t}
\end{aligned} \tag{1}
$$

where subscript i denotes the level of analysis of *individual page visits*, r denotes NUTS-3 *regions*, l denotes *language versions* of the app, and *t* denotes the *time* in days. In total, the 105 regions provide 644 region-language *platform versions*.

Table 1 lists all variables. The dependent variable $TopicView_i$ is set to missing whenever $TopicSupply_{r,l,t} = 0$ to restrict the analysis to visits where the topic could, in principle, have been viewed. This design yields topic-specific samples from 885,103 (Racism & Discrimination) to 2,928,689 (Work & Training); see Table A4. $TopicSupply_{r,l,t}$ captures the supply side by measuring the share of topic-specific URLs within the platform version observed in i, i.e., the emphasis a region places on that topic. $Lang_l$ groups display languages into seven categories to obtain sufficient cases for differentiation (RQ3), thereby reflecting major immigrant communities in Germany.

Table 1: Variables used in the regression models and their operationalization.

Variable	Symbol	Type	Scale / Coding
Topic view (DV)	$TopicView_i$	Binary	$1 =$ page contains the topic; $0 =$ otherwise; **missing** if $TopicSupply_{r,l,t} = 0$
Topic supply (share of URLs)	$TopicSupply_{r,l,t}$	Continuous (percentage)	0-100; $(\frac{Nurl(topic)_{r,l,t}}{Nurl(overall)_{r,l,t}}) \times 100$

Variable	Symbol	Type	Scale / Coding
Language group	$Lang_l$	Categorical (7 groups)	Levels: (1) English; (2) Arabic; (3) Farsi; (4) Russian/ Ukrainian; (5) Turkish/ Kurdish; (6) East- European languages;[13] (7) African languages[14]
Month indicator	$Month_i$	Categorical (factor)	Monthly fixed effects
Weekend	$Weekend_i$	Binary	$1 =$ weekend day; $0 =$ weekday
Referred from social media	$SocialMedia_i$	Binary	$1 =$ referred from social media; $0 =$ otherwise
App size (URLs)	$AppSize_{r,l,t}$	Continuous (count)	Absolute number of distinct URLs
Time since launch	$MonthsLive_{r,l,t}$	Continuous (duration)	Duration since launch of (r, l) platform version
District indicator	$District_r$	Categorical	District fixed effects

Monthly indicators $Month_i$ absorb shocks or usage spikes (e.g., after the influx of refugees from Ukraine). $Weekend_i$ accounts for potential residual activity by administration staff assumed to occur on workdays (see Table A3). $SocialMedia_i$ flags social-media referrals, e.g., to control for advertisement-driven over-representation of topics. $AppSize_{r,l,t}$ (distinct URL count) proxies invested human resources into the app. $MonthsLive_{r,l,t}$ (time since launch) allows topic emphasis to evolve with experience and inter-agency exchange. Finally, $District_r$ introduces district fixed effects to absorb remaining unobserved regional heterogeneity.

Standard errors in Equation 1 are clustered at the level of platform versions r, l, as the web-

[13] Romanian (N = 333158), Bulgarian (N = 107909), Croatian (N = 98458), Polish (N = 67152), Hungarian (N = 14213), Serbian (N = 8068), Albanian (N = 2721), Czech (N = 217).

[14] French (N = 35327), Amharic (N = 10295), Tigrinya (N = 1594), Somali (N = 101).

page usage is likely more similar within these groups, for instance due to varying quality and supply of information across topics as well as varying interest by language.

Based on Equation 1, predictive margins of the models are calculated using STATA's margin command (StataCorp, 2023) to map page visit probabilities for different specifications in $TopicSupply_{r,l,t}$: First, for analyzing RQ2 (*alignment with user engagement*), predictive margins are calculated at the sample's median of $TopicSupply_{r,l,t}$ across i to obtain insights based on a simplified setting in which individual page visit probabilities are only compared to the most "typical" topic shares. In a second step, this restriction is eased and predictive margins are calculated at five equally spaced supply values in $TopicSupply_{r,l,t}$. The minimum and maximum values of these five values are defined by each topic's observed distribution across i to ensure that they are sufficiently covered by the data.[15] The idea here is to use the observed variation in $TopicSupply_{r,l,t}$ across platform versions and compare page visit probabilities across multiple realistic supply levels. Finally, for analyzing RQ3 (*language-specific variations*), predictive margins are calculated at the same sample's median of $TopicSupply_{r,l,t}$ as before and by categorical language levels defined in $Lang_l$, to assess if page visit probabilities for typical supply levels vary by language groups.

Findings

Topic coverage (RQ1)

Figure 4 shows information on how often and deeply administrations offer the 20 topics considered. Panel (a) displays the share of platform versions providing at least one page URL for which a topic has been identified. Some topics are served by almost all platform versions, such as *Work & Training*, which appears in 91.2% of platform versions across time, just like *Counseling & Help* (89.2%), *Learning German* (87.5%), or *Family & Children* (85.7%). Much less frequently offered is information on *Opening Hours & Holidays* (29.4%), *The Elderly* (15.0%), or *Racism & Discrimination* (13.3%). A topic that experienced major changes in coverage comprises information on *COVID-19*, which was covered in 90.1% of platform versions in November 2021 and decreased to 9.7% in February 2024.

As a measure of relative topic depth, panel (b) reports distributions of topic shares in percent, given that they are provided by at least one page (URL >= 1), and on the level of platform-version-days. For instance, across all platform-version-days for which *Work & Training* is observed with at least one

[15] Due to the right-skewed distribution of the topic supply, the start and end values are defined by each topic's 1st and 90th percentiles of $TopicSupply_{r,l,t}$.

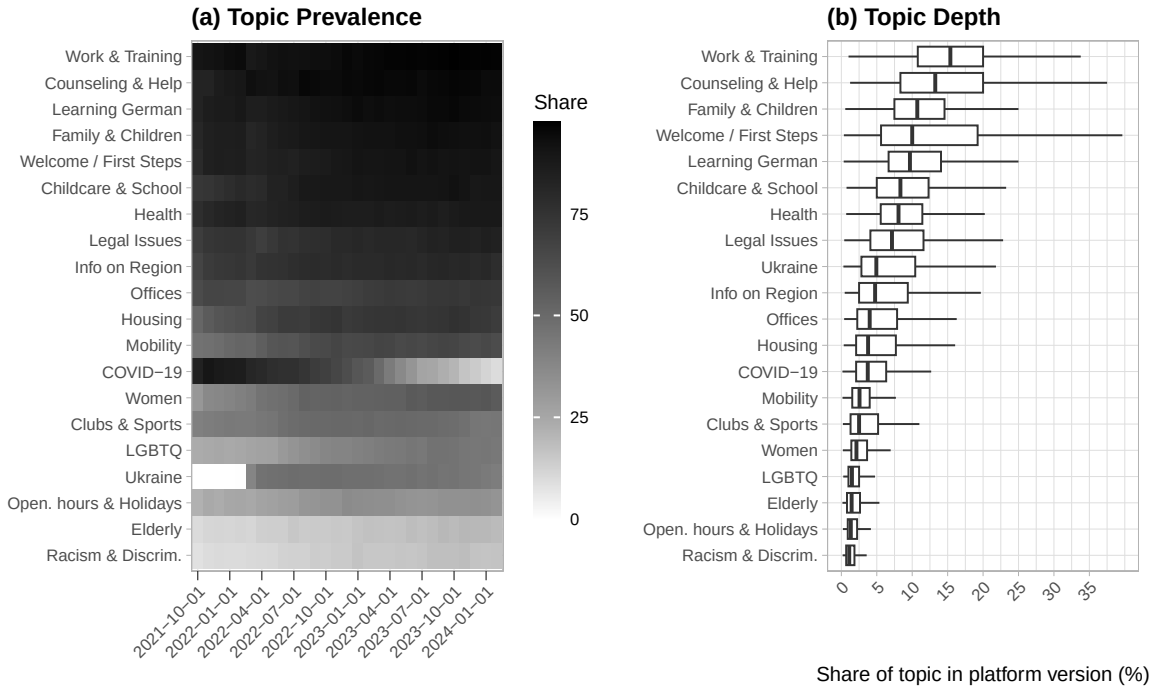

Figure 4. Prevalence and depth in integration topic: (a) Share of region-language platform versions that contain at least one URL for the given topic, by month. (b) Relative topic depth, defined for platform versions as the number of URLs in the topic divided by the total number of URLs, measured daily. Box plots summarize the distribution (min, Q25, median, Q75, max) of relative topic depth across platform-version-days with at least one URL in the respective topic (N = 2,275-15,444 platform-version-days, depending on topic).

page, the median coverage this topic consumes of its overall platform version is 15.4%. In general, the topics that are offered most often also tend to be the ones with the highest shares. Nevertheless, there is major variation observable within topics and across platform versions. For example, considering *Work & Training*, the upper quarter of platform versions dedicates a share of 20% or more to this topic, while the lower quarter dedicates only 11.8% or less. Even more evident, *Welcome & First Steps* shows ≤8.3% in Q1 versus ≥20.0% in Q3.

Overall, these findings indicate that while certain topics-particularly those targeting the acquisition of resources and language-are consistently represented, there is substantial variation across platform versions in terms of frequency and intensity of topic provision, indicating that administrations may be selectively emphasizing specific themes. The subsequent analyses further examine how these varying emphases are associated with actual user engagement.

Probability of page visits at median supply (RQ2)

Turning to the multivariate models assessing the alignment between topic supply and user engagement, Figure 5 presents two panels. Panel (a) illustrates the probability of page visits (represented by point markers) computed at the sample's median values of $TopicSupply_{r,l,t}$ (indicated by diamond markers). One way to think about the relationship between the two is that if users were to interact indiscriminately with all topic contents ("as if random"), page visit probabilities (in %) would equal $TopicSupply_{r,l,t}$ (in %). Consequently, any deviation between these measures indicates under- or over-engagement relative to the topic supply. Panel (b) explicitly plots these discrepancies by showing the percentage point differences obtained by subtracting the median topic supply from the probability of page visits, with a reference line denoting perfect alignment.

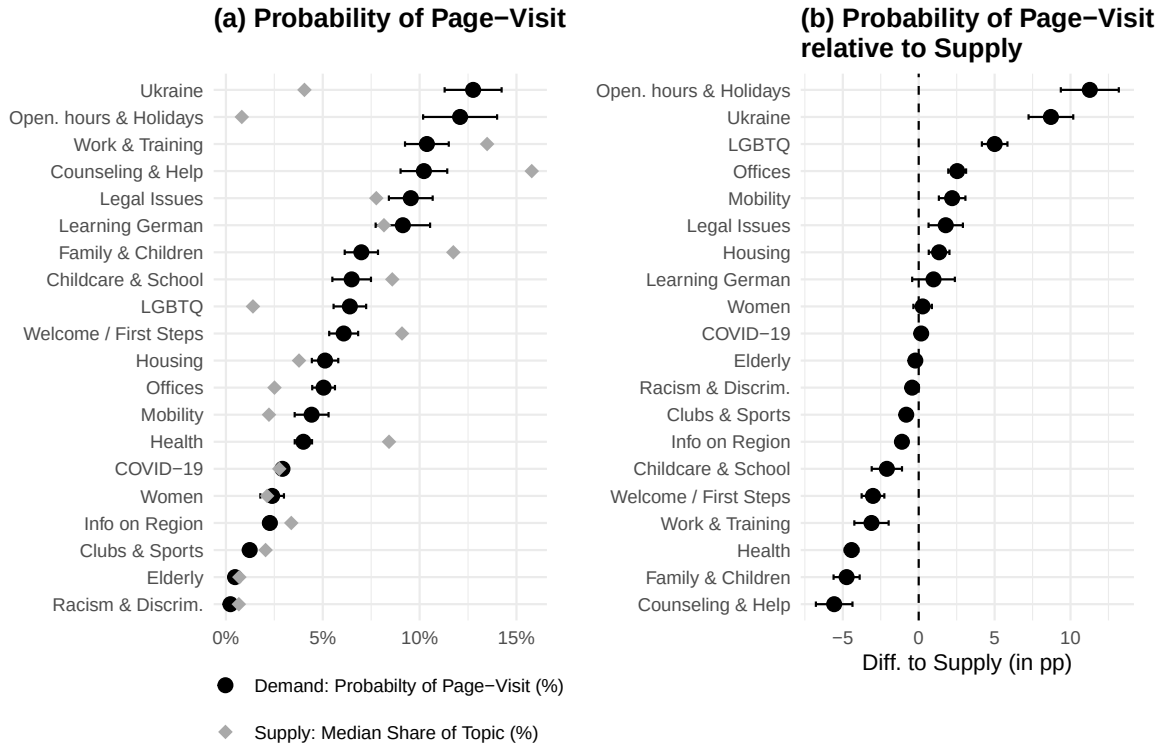

Figure 5. Prediction of integration-platform visits at median supply: Panel (a): probability of a page visit for each topic in % (with 95% CIs), evaluated at the sample's median of $TopicSupply_{r,l,t}$ in %, marked as diamonds. Panel (b): percentage-point differences (probability minus median supply); vertical dashed line at 0 denotes perfect alignment.

One topic that stands out is *Opening Hours & Holidays* with an estimated 12.1% probability of

a page visit. With a median supply of only 0.8% of URLs, it represents a very large positive discrepancy between page visits and supply (+11.3pp). Due to its federal structure, school holidays and many public holidays vary from state to state in Germany, and opening hours of stores may additionally vary across districts or municipalities. Thus, the topic represents a type of information that is of central importance for users living in the region to organize their everyday lives, and is a type of information not often available in multiple languages outside of the app due to the small-scale particularities. Furthermore, more than 70% of region-language versions do not cover this topic at all, which may indicate a lack of awareness on behalf of the responsible administrations around the importance of offering this topic.

The topic with the highest likelihood of a page visit in general is the information for refugees from *Ukraine*, at 12.8%. This is particularly noteworthy, as the median supply for this topic is only 4.1%, i.e., the relative interest lies substantially above the relative provision (+8.7pp). Looking at the specific URLs contributing to this discrepancy, it is especially those that specifically address refugees from Ukraine on the start screen. Thus, the first URL-level that defines the selection of tiles frequently reads “Help to people from Ukraine”, “Information for people from Ukraine” or “Ukraine crisis”. Nevertheless, many pages also represent forwarding pages from the ones above, i.e., providing more detailed information following a click on the first tile. They contain information on aspects such as driving licenses, accommodation, or information on social benefits in Germany (ibid.). Overall, this implies that directly addressing Ukrainians on the start screen attracted relatively high interest from users compared to other selections that were topic-specific only.

Information specific to people who are lesbian, gay, bisexual, transgender, or queer (*LGBTQ*) is visited with a probability of 6.4%. Given the supply share of only 1.4%, this topic likewise represents information with relatively high engagement (+5.0pp). The language versions driving this interest are first and foremost Farsi and Dari, and to a lesser extent Arabic (Figure A4). Farsi, along with its dialect Dari, is most prevalent in Iran and Afghanistan (where it is called Dari). In these two countries, LGBTQ individuals endure severe persecution and discrimination, with same-sex relations being criminalized and punishable by violence, imprisonment, or even death (Ahmady, 2018; Razai et al., 2023). In this sense, the comparatively high interest may reflect the specific causes of migration for these specific groups of refugees and their associated demand for specific information.

Some topics are accessed with a probability approximately matching their median supply. These include, among others, *Legal Issues* (+1.8pp), *Learning German* (+1.0pp), *Housing* (+1.3pp), information on *COVID-19* (+0.2pp), and information specific to *Women* (+0.3pp). The assessment by the re-

gional administrations regarding the approximate relevance of these topics can therefore be considered relatively accurate. This is particularly noteworthy for *Learning German* and *Legal issues*, as these topics are given a relatively high proportion in the application. Topics whose page visit probabilities are remarkably lower than their shares of supply include *Counseling & Help* (-5.6pp), *Family & Children* (-4.7pp), *Health* (-4.4pp), and *Work & Training* (-3.1pp).

Probability of page visits by supply share (RQ2)

Figure 6 shows predicted page-visit probabilities across five levels of $TopicSupply_{r,l,t}$; values above (below) the 45° line indicate demand exceeding (falling short of) supply. Some findings should be highlighted. First, the increased demand for topics such as *Opening Hours & Holidays*, *Ukraine* and *LGBTQ* identified for the median supply before, remains consistent across other observed supply ranges. This is indicated by the predicted page visits running relatively parallel to the diagonal, suggesting stable positive discrepancies between supply and demand. Second, all topics show a positive relationship between supply (the share of topics provided on the platform) and demand (the probability of page visits). As the supply of content on a particular topic increases, the probability of page visits generally rises, but at varying rates depending on the topic. Third, certain topics, such as *Housing*, *Legal Issues*, *Learning German*, *LGBTQ*, and *Opening Hours & Holidays*, can be described as both highly demanded and responsive. These topics have demand curves that are equal to or even above the supply line, and their slopes are at least as steep as the diagonal. This suggests that (a) there is higher user engagement as more content is provided, and (b) there is no observable saturation effect, as the probability of page visits does not flatten with increasing supply levels. Fourth, topics such as *Mobility* and *Offices* are best described as highly demanded but less responsive. These topics show predicted page visits above the diagonal throughout the observed supply range, but their demand curves are flatter than the diagonal, indicating that further increases in supply could eventually result in excess supply relative to demand. Fifth, topics such as *Clubs & Sports*, *COVID-19*, and *Information on the Region* may be best described as moderately in demand but low in responsiveness. These topics show predicted page visits that are equal to or above the supply line at lower supply levels, but demand does not proportionally increase with additional content. Consequently, their page visit curves are relatively flat and fall below the supply diagonal. Strictly, topics such as *Childcare & School* and *Welcome / First Steps* also fall into this category. However, these are still relatively more responsive, as their predicted page visits remain just below the diagonal.

Finally, the topics *Racism & Discrimination* and *The Elderly* exhibit both low demand and low

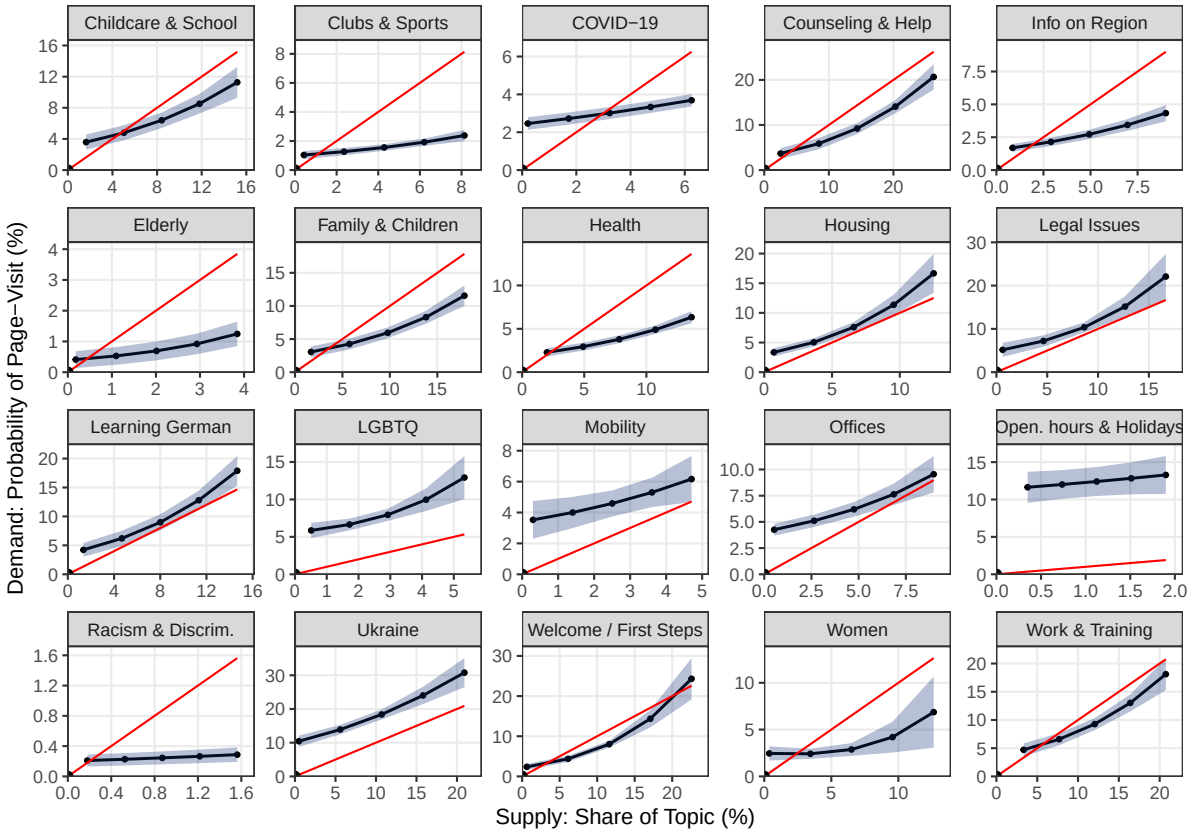

Figure 6. Prediction of integration-platform visits over varying supply levels: Predictive margins are evaluated at five equally spaced values of $TopicSupply_{r,l,t}$ for each topic panel. The red 45° line indicates perfect alignment-where the predicted probability of a page visit (%) equals the share of information provision (%). Curves above the line imply engagement exceeds supply; curves below imply engagement falls short relative to supply.

responsiveness. The demand curves for these topics are significantly below the supply line and nearly flat, indicating that increases in content are not associated with a substantial increase in visits within the supply ranges considered.

Probability of page visits by language (RQ3)

Finally, this section addresses RQ3, which investigates whether information needs vary by language – reflecting the diverse socio-economic and cultural backgrounds of the immigrant population in Germany (German Federal Office for Statistics, 2024a). Figure 7 displays, for selected topics and language versions, the difference between page visit probability and the median topic supply. All topic and language

combinations are provided in the supplemental materials (see Figure A4).

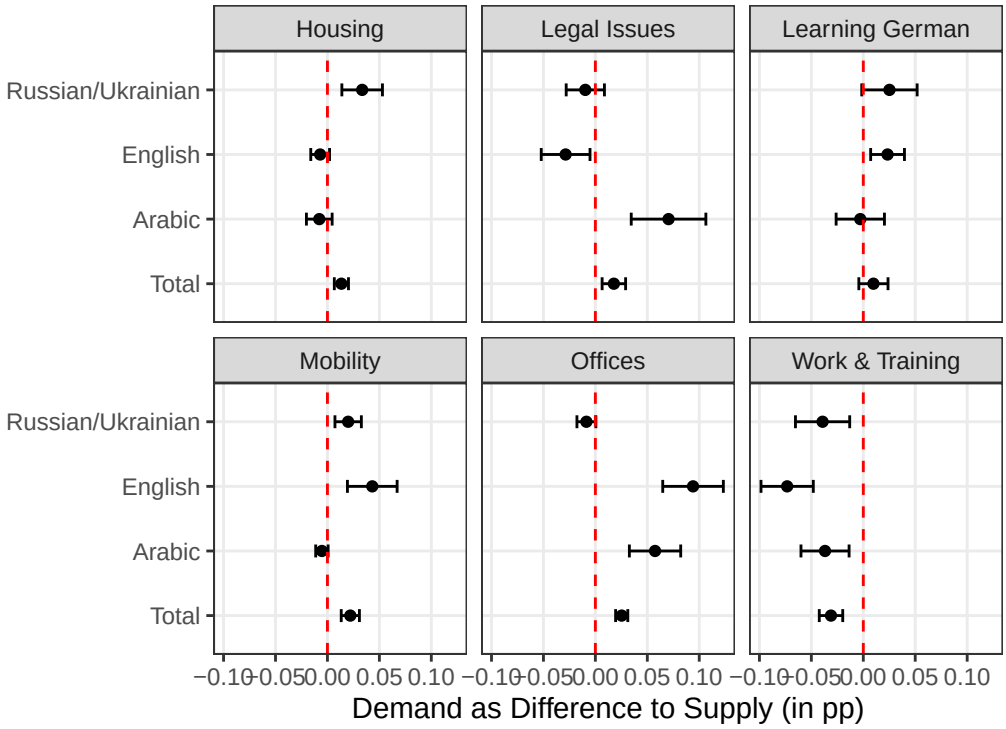

Figure 7. Prediction of integration-platform visits at median supply, by language: Percentage-point differences (probability of page visits minus median supply) by used language; vertical dashed line at 0 denotes perfect alignment.

The engagement with the topic area of *Housing* is comparatively high for users in Ukrainian and Russian. One plausible explanation is the relatively short duration of stay of this group during the observation period (German Federal Office for Statistics, 2024b). Many Ukrainian refugees were just about to move from reception centers into private flats, compared to, for instance, Arabic-speaking Syrians, who on average had been living in Germany for several years longer (German Federal Office for Statistics, 2024b; Siegert et al., 2023). Similarly, *Learning German* shows higher engagement in Ukrainian and Russian platform versions. Due to their relatively short duration of stay and long waiting times for language courses, many Ukrainian refugees are still in the early stages of integration and actively seek information on language learning, whereas other groups may have already completed their courses prior to the observation period (Kosyakova et al., 2024).

Likewise, *Legal Issues* show higher engagement among Arabic-speaking users, likely reflecting

their often complex legal situations during the observation period and the associated increased need for information. Thus, while many Syrians who arrived in 2015 and 2016 are eligible to apply for naturalization due to sufficiently long periods of residence (Kassam & Becker, 2023), the ones who arrived later may strive obtaining a permanent residence permit. Furthermore, some Arabic-speaking groups like Iraqis or Lebanese generally face lower protection rates and thus more uncertain legal statuses, which can further drive the demand for legal information (German Federal Office for Statistics, 2025).

Other noticeable findings include a comparatively high level of engagement from English-speaking users in the topics of mobility and public offices, as well as a reduced demand for information on work and training compared to other languages considered. It is important to note that the English version likely attracts a heterogeneous mix of immigrant groups-often serving as a fallback when users' native languages are unavailable or poorly covered-making the interpretation of these results particularly challenging.

Overall, these patterns suggest that engagement with integration-related topics varies markedly by language, reflecting distinct integration trajectories and legal circumstances across immigrant groups.

Conclusion and discussion

Multilingual information helps refugees and other immigrants navigate their first steps; digital channels are essential because many obtain information via smartphones, tablets, or PCs (Borkert et al., 2018; Jauhiainen et al., 2021; Merisalo & Jauhiainen, 2019; Stapf, 2019). In federal systems like Germany, centralized services risk low local relevance, while city and district administrations often lack the IT and staff to prepare and update content across languages (Kühn & Ziegler, 2024; Tür an Tür - Digitalfabrik gGmbH, 2024).

This study examines a digital platform that provides municipalities a shared infrastructure to publish locally tailored, multilingual information via app and web. Because content creation is decentralized, many administrations report resource constraints (Tür an Tür - Digitalfabrik gGmbH, 2024). The analysis therefore investigates which topics are prioritized under such constraints and how these priorities correspond to users' engagement. Using up to 2.9 million visit records, multivariate logistic regressions are estimated to model engagement, complemented by evidence on regional diffusion and language availability.

The study arrives at several key findings. First, thematic coverage differs markedly across

districts and cities: labor-market integration and language acquisition are almost always available, while mobility, opening hours, and anti-racism are less common. Even when topics are present, their share of total content varies widely, indicating that districts and cities may prioritize different topics or make assumptions about which topics are likely relevant to users within their jurisdictions.

Second, the relative supply of a topic is positively associated with visit probability, with strength varying by topic and access language. Immediate, practical information (e.g., opening hours/holidays, office addresses, mobility/tickets) and audience-specific pages (e.g., for Ukrainian refugees or LGBTQ users) attract disproportionate engagement, yet are unevenly provided, indicating a potential undersupply in some areas.

Third, legal status, language, and housing show consistently high demand and remain responsive as content expands: up to ~10-15% of a local platform, we find no saturation in visit probabilities. This suggests that users tend to explore more detailed and extensive subpages when such content is available. Regions that currently provide limited information on these topics may enhance their target audience's knowledge by expanding the content offered. This is particularly relevant because these topics primarily concern the acquisition of essential resources (Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration, 2021) that are crucial in the early stages after arrival and form the basis for many subsequent integration processes.

Fourth, it was demonstrated that the relative demand for certain topics varies depending on the access language, and these differences can often be plausibly linked to the situations commonly experienced by these groups during the observation period. For example, the demand for information on housing is particularly high among Russian- and Ukrainian-speaking users, who often still live in communal accommodations, while the demand for legal topics is higher among Arabic-speaking users, whose legal situations tend to be relatively complex.

Some important limitations remain. First, engagement of users is not only shaped by their interest and needs but also by the app's design. Although all regions use the same uniform platform, they retain some design discretion that can affect engagement. For example, the arrangement of home-screen tiles may vary and influence user behavior. An implicit assumption in the analyses is that, on average, the number of pages on a topic is linearly related to its in-app visibility. If, at equal topic depth, certain topics are systematically placed more prominently, part of the observed engagement may be attributable to this placement.

Second, the presented supply-demand comparison is designed to show how regions' relative thematic priorities are associated with users' relative likelihood of access across core topic areas. When a topic receives strong emphasis alongside similarly high engagement, its relevance is supported. Conversely however, low supply with high engagement does not necessarily imply undersupply if the topic can be covered adequately on a few pages (e.g., holiday dates and opening hours). Interpretation on relevance therefore depends on combining these results with earlier findings indicating whether these topics are offered at all (Figure 4). Furthermore – while the analysis of engagement across different supply intensities approximates the identification of an "optimal" relative emphasis per topic (Figure 6) – this perspective is also constrained by the variation that ca be observed across platform versions. For example, it remains unknown whether engagement in "Learning German" would continue to rise proportionally if supply shares exceeded the observed values (~15%). Therefore, the "optimal" topic emphasis may be unobserved.

The practical implications of these findings center around optimizing the communication between local administrations and immigrants within their jurisdiction. With limited resources, some regions likely have made simplified assumptions about information needs. The results provide initial insights into the topics of greatest importance and can guide informed decisions beyond this specific use case. Moreover, a closer look at regional demographics (e.g., length of stay, country of origin) or surveys of immigrants could more effectively address diverse information requirements. Ensuring adequate funding for administrations is essential to enable such tailored strategies and enhance the overall utility of local digital platforms for refugees.

Given the previously unknown data, it is also one central aim of this work to provide the descriptive knowledge necessary to form a basis for further ideas and follow-up studies. One promising avenue involves examining local political contexts. Prior research has linked regional asylum recognition rates, court decisions on asylum law, and expulsion rates to residents' or local governments' attitudes toward immigration (Gundacker et al., 2024; Schneider et al., 2020). These political peculiarities may also be associated with local communication on integration, with administrations favoring pro-integration policies potentially emphasizing other topics (e.g., social services, education, or legal support), while other administrations might prioritize rapid labor market integration or security.

Another promising direction leverages the platform's staggered roll-out across regions. This setup is well-suited to inferential methods-such as staggered difference-in-differences-to estimate the causal impact of the platform's introduction (Callaway & Sant'Anna, 2021; Goodman-Bacon, 2018).

By comparing regions before and after implementation, as well as against areas yet to adopt the platform, researchers may assess effects on region-level outcomes (e.g., refugee employment or language course participation) or individual-level outcomes from panel surveys such as the IAB-BAMF-SOEP study on refugees (Brücker et al., 2024). This might enable an explicit measurement of outcomes, related to navigational integration (Harder et al., 2018) or emotional integration and institutional trust (de Alpuim-Gonçalves et al., 2025; Potter et al., 2024). Similarly, the continuous updates and maintenance of platform versions by local administrations represent quasi-experimental interventions to causally identify how different content updates affect user engagement compared to regions that do not implement such updates simultaneously. Complementing the limited number of experimental studies on multilingual interventions (e.g., Kiropoulos et al., 2011; Tjaden et al., 2022) with such observational approaches may represent a promising avenue for future research and would directly address some of the limitations outlined above.

Acknowledgements

I acknowledge the constructive feedback on earlier drafts of this work from participants of the 2024 online writing workshop at the WZB Berlin Social Science Center. Furthermore, I want to acknowledge support from Moritz Marbach for feedback during early idea finding and Niklas Harder for later feedback.

Funding and competing interests

The author declares no conflict of interest. Specifically, the author declares no affiliation with the data provider of this study. The author received no financial support for the research or publication of this article.

Bibliography

Ahmady, K. (2018). Migration and Gender for Iranian LGBT. *ScholarWorks@Arcadia*, *4*(1), 2. https://scholarworks.arcadia.edu/agsjournal/vol4/iss1/2/?utm_source=scholarworks.arcadia . edu % 2Fagsjournal % 2Fvol4 % 2Fiss1 % 2F2&utm _ medium = PDF&utm _ campaign=PDFCoverPages

Aumüller, J. (2017, September). Die kommunale Integration von Flüchtlingen. In *Handbuch Lokale Integrationspolitik* (pp. 173–198). Springer VS. https://doi.org/10.1007/978-3-658-13409-9_8

BAMF. (2024). Bericht zur Integrationskursgeschäftsstatistik 2023. https : / / www . bamf . de / SharedDocs / Anlagen / DE / Statistik / Integrationskurszahlen / Bundesweit / 2023 - integrationskursgeschaeftsstatistik-gesamt_bund.pdf?__blob=publicationFile&v=6

Barreto, C., Berbée, P., Gallegos Torres, K., Lange, M., & Sommerfeld, K. (2022). The civic engagement and social integration of refugees in germany. *Nonprofit Policy Forum*, *13*(2), 161–174. https://doi.org/10.1515/npf-2022-0015

Battisti, M., Peri, G., & Romiti, A. (2022). Dynamic effects of co-ethnic networks on immigrants' economic success. *The Economic Journal*, *132*(641), 58–88. https://doi.org/10.1093/ej/ueab036

Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration. (2021). Integration in Deutschland: Erster Bericht zum indikatorgestützten Integrationsmonitoring. https : / / www . integrationsbeauftragte . de / resource / blob / 1872554 / 1881534 / 91dd9a36d492ca4ba29482072473fa62/bericht-integrationsmonitoring-data.pdf

Borkert, M., Fisher, K. E., & Yafi, E. (2018). The Best, the Worst, and the Hardest to Find: How People, Mobiles, and Social Media Connect Migrants In(to) Europe. *Social Media + Society*, *4*(1), 2056305118764428. https://doi.org/10.1177/2056305118764428

Bradley, L., Lindstrm, N. B., & Hashemi, S. S. (2017). Integration and Language Learning of Newly Arrived Migrants Using Mobile Technology. *Journal of Interactive Media in Education*, *2017*(1). https://doi.org/10.5334/jime.434

Brücker, H., Kosyakova, Y., Goebel, J., Grabka, M. M., Rother, N., Schröder, C., Zinn, S., Bartels, C., Dräger, J., Eckert, J., Franken, A., Gauer, M., Gerike, M., Griese, F., Beckmannshagen, M., Kara, S., Krause, P., Liebau, E., Maddox, A., … Deutsches Institut Für Wirtschaftsforschung (DIW Berlin). (2024). IAB-BAMF-SOEP-Befragung Geflüchteter 2022. https://doi.org/10.5684/SOEP.IAB-BAMF-SOEP-MIG.2022

Bundesamt für Migration und Flüchtlinge. (2022). Erstverteilung der Asylsuchenden (EASY). https://www.bamf.de/DE/Themen/AsylFluechtlingsschutz/AblaufAsylverfahrens/Erstverteilung/erstverteilung-node.html

Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR). (2024). Laufende Raumbeobachtung des BBSR - INKAR, Ausgabe 03/2024. https://www.inkar.de

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Cesare, N., Lee, H., McCormick, T., Spiro, E., & Zagheni, E. (2018). Promises and pitfalls of using digital traces for demographic research. *Demography*, *55*(5), 1979–1999. https://doi.org/10.1007/s13524-018-0715-2

Commissioner for Migration, Refugees and Integration (IB). (2024, December 11). *Teilhabe in der Einwanderungsgesellschaft: Ein wissenschaftsbasierter und indikatorengestützter Lagebericht zum Stand der Integration in Deutschland – 14. Integrationsbericht der Beauftragten der Bundesregierung für Migration, Flüchtlinge und Integration: Lagebericht integrationsbericht* [Brochure, Item number 44436]. Die Beauftragte der Bundesregierung für Migration, Flüchtlinge und Integration. https://www.publikationen-bundesregierung.de/resource/blob/2277604/2324696/c9c424884ff0c8c109c13aae3e8f0d3a/14-lagebericht-integration-download-ba-ib-data.pdf?download=1

de Alpuim-Gonçalves, S. M. B., Garrido, A. S. M. F. C., Garrigos, G. C., Ferré-Grau, C., Sousa, S. C. S., & da Cruz Sequeira, C. A. (2025). Interventions Promoting Positive Mental Health for Migrant and Refugee Adults in the Community: A Scoping Review. *Int. Migration &. Integration*, 1–37. https://doi.org/10.1007/s12134-024-01223-z

El-Desouky, M. (2023, April 25). Gtranslate: Translate between different languages. https://doi.org/10.32614/cran.package.gtranslate

Eule, T. G. (2017, February 8). Ausländerbehörden im dynamischen Feld der Migrationssteuerung. transcript Verlag. https://doi.org/10.14361/9783839433324-007

Federal Ministry for Economic Affairs and Climate Action. (2024). Gleichwertigkeitsbericht 2024: Für starke und lebenswerte Regionen in Deutschland. https://www.bmi.bund.de/SharedDocs/downloads/DE/veroeffentlichungen/2024/gleichwertigkeitsbericht.pdf?__blob=publicationFile&v=6

Federal Ministry of the Interior and Community. (2019, July). Unser plan für deutschland. gleichwertige lebensverhältnisse überall. https://www.bmfsfj.de/resource/blob/137240/e94cf2ffab8768fd37a1e632db3ee51e/schlussfolgerungen-kommission-gleichwertige-lebensverhaeltnisse-langversion-data.pdf

Filsinger, D., & Gesemann, F. (2017). Kommunale Integrationskonzepte: Verbreitung, Funktion und Wirksamkeit. *Wissenschaftliches Gutachten im Auftrag des Sachverständigenrats deutscher Stiftungen für Integration und Migration (SVR)*. https://www.altenburgerland.de/media/full/3733/Kommunale-Integrationskonzepte-Verbreitung-Funktion-und-Wirksamkeit.pdf

Fink, P., Hennicke, M., & Tiemann, H. (2019). *Ungleiches Deutschland: Sozioökonomischer Disparitätenbericht 2019*. Friedrich-Ebert-Stiftung. http://www.fes.de/cgi-bin/gbv.cgi?id=15400&ty=pdf

German Federal Office for Statistics. (2024a, September). Bevölkerung mit Migrationshintergrund [Online; accessed 8. Sep. 2024]. https : / / www . destatis . de / DE / Themen / Gesellschaft - Umwelt / Bevoelkerung / Migration - Integration / Publikationen / _publikationen-innen-migrationshintergrund.html

German Federal Office for Statistics. (2024b, September). Durchschnittliche Aufenthaltsdauer der Ausländer: Bundesländer, Stichtag, Geschlecht, Ländergruppierungen/Staatsangehörigkeit [Online; accessed 8. Sep. 2024]. https : / / www - genesis . destatis . de / genesis / / online ? operation = table&code = 12521 - 0031&bypass = true&levelindex=1&levelid=1725807438669

German Federal Office for Statistics. (2025, February). Schutzsuchende nach Schutzstatus, Regionen und Herkunftsländern [[Online; accessed 3. Feb. 2025]]. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/Tabellen/schutzsuchende-staatsangehoerigkeit-schutzstatus.html

Goebel, J., Grabka, M. M., Liebig, S., Schröder, C., Zinn, S., Bartels, C., Franken, A., Gerike, M., Griese, F., Halbmeier, C., Kara, S., Krause, P., Liebau, E., Nebelin, J., Petrenz, M., Satilmis, S., Siegers, R., Steinhauer, H. W., Süttmann, F., … Deutsches Institut Für Wirtschaftsforschung (DIW Berlin). (2023). Sozio-oekonomisches Panel, Daten der Jahre 1984-2021 (SOEP-Core, v38.1, EU Edition). https://doi.org/10.5684/SOEP.CORE.V38.1EU

Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. https://doi.org/10.3386/w25018

Grün, B., & Hornik, K. (2024). *Topicmodels: Topic models* [R package version 0.2-17]. https://doi.org/10.32614/CRAN.package.topicmodels

Gundacker, L., Kosyakova, Y., & Schneider, G. (2024). How regional attitudes towards immigration shape the chance to obtain asylum: Evidence from germany. *Migration Studies*, mnae002. https://doi.org/10.1093/migration/mnae002

Harder, N., Figueroa, L., Gillum, R. M., Hangartner, D., Laitin, D. D., & Hainmueller, J. (2018). Multidimensional measure of immigrant integration. *Proceedings of the National Academy of Sciences*, *115*(45), 11483–11488. https://doi.org/10.1073/pnas.1808793115

IT Planning Council. (2017, March). Mehrsprachigkeit in modernen Bürgerdiensten [Online; accessed 8. Sep. 2024]. https://www.it-planungsrat.de/beschluss/beschluss-2017-17

Jauhiainen, J. S., zrmez, S., & Tursun, Ö. (2021). Internet and social media uses, digital divides, and digitally mediated transnationalism in forced migration: Syrians in Turkey. *Global Networks*, *22*(2), 197–210. https://doi.org/10.1111/glob.12339

Jurić, T. (2022). Ukrainian refugee integration and flows analysis with an approach of big data: Social media insights. *medRxiv*. https://doi.org/10.1101/2022.04.18.22273958

Kanas, A., Kosyakova, Y., & Vallizadeh, E. (2022). Linguistic enclaves, sorting, and language skills of immigrants. *Journal of Immigrant & Refugee Studies*, *22*(4), 847–861. https://doi.org/10.1080/15562948.2022.2132572

Kassam, K., & Becker, M. (2023). Syrians of today, Germans of tomorrow: the effect of initial placement on the political interest of Syrian refugees in Germany. *Front. Polit. Sci.*, *5*, 1100446. https://doi.org/10.3389/fpos.2023.1100446

Khalil, S., Kohler, U., & Tjaden, J. (2022). Is there a rural penalty in language acquisition? evidence from germany's refugee allocation policy. *Frontiers in Sociology*, *7*. https://doi.org/10.3389/fsoc.2022.841775

Khalil, S., & Tjaden, J. (2025). How reception centers affect the integration of asylum seekers and recognized refugees. *European Societies*, 1–32. https://doi.org/10.1162/euso_a_00018

Kiropoulos, L. A., Griffiths, K. M., & Blashki, G. (2011). Effects of a multilingual information website intervention on the levels of depression literacy and depression-related stigma in greek-born and italian-born immigrants living in australia: A randomized controlled trial. *J Med Internet Res*, *13*(2), e34. https://doi.org/10.2196/jmir.1527

Kosyakova, Y., Gatskova, K., Koch, T., Adunts, D., Braunfels, J., Goßner, L., Konle-Seidl, R., Schwanhäuser, S., & Vandenhirtz, M. (2024). Arbeitsmarktintegration ukrainischer Geflüchteter: Eine internationale Perspektive. *IAB-Forschungsbericht*, 2024–16. https://doi.org/10.48720/IAB.FB.2416

Kosyakova, Y., Kristen, C., & Spörlein, C. (2021). The dynamics of recent refugees' language acquisition: How do their pathways compare to those of other new immigrants? *Journal of Ethnic and Migration Studies*, *48*(5), 989–1012. https://doi.org/10.1080/1369183x.2021.1988845

Kuhlmann, S., Oehlert, F., & Reusch, M. C. (2024). Verwaltungsverflechtungen in der Flüchtlingsintegration – ein Vergleich zwischen Deutschland, Schweden und Frankreich. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-43195-2_55-1

Kühn, B., & Ziegler, F. (2024). Weiter am Limit? Zur Lage der Kommunen bei der Aufnahme Geflüchteter. https : / / mediendienst - integration . de / fileadmin / Dateien / Expertise _ Kuehn_Ziegler_Umfrage_Kommunen_Mai_2024.pdf

Kukulska-Hulme, A., Gaved, M., Paletta, L., Scanlon, E., Jones, A., & Brasher, A. (2015). Mobile Incidental Learning to Support the Inclusion of Recent Immigrants [Online; accessed 6. Sep. 2024]. https://doi.org/10.18848/1835-9795/CGP/v07i02/58070

Liebau, E., & Romiti, A. (2014). *Migranten investieren in Sprache und Bildung* (tech. rep. No. 43). Berlin, Deutsches Institut für Wirtschaftsforschung (DIW). https://hdl.handle.net/10419/104030

Lueders, H. (2023). Not always backlash: The rise of refugee support organizations during the european refugee crisis. *Available at SSRN 4261074*. https://doi.org/10.2139/ssrn.4261074

Marbach, M., Vallizadeh, E., Harder, N., Hangartner, D., & Hainmueller, J. (2025). Does ad hoc language training improve the economic integration of refugees? evidence from germany's response to the syrian refugee crisis. *Journal of the Royal Statistical Society Series A: Statistics in Society*. https://doi.org/10.1093/jrsssa/qnae106

Mediendienst Integration. (2025, January). Wie geht es mit den Integrationskursen weiter? [Online; accessed 28. Jan. 2025]. https://mediendienst-integration.de/artikel/wie-geht-es-mit-den-integrationskursen-weiter.html

Merisalo, M., & Jauhiainen, J. S. (2019). Digital Divides among Asylum-Related Migrants: Comparing Internet Use and Smartphone Ownership. *Tijdschrift voor economische en sociale geografie*, *111*(5), 689–704. https://doi.org/10.1111/tesg.12397

Ostermayer, M., Hennicke, M., & Tiemann, H. (2022, December 5). Ungleiches Deutschland: Regionale Disparitäten und was dagegen getan werden kann. De Gruyter. https://doi.org/10.1515/9783110701678-001

Phan, X.-H., Nguyen, L.-M., & Horiguchi, S. (2008). Learning to classify short and sparse text & web with hidden topics from large-scale data collections. *Proceedings of the 17th international conference on World Wide Web*, 91–100. https://doi.org/10.1145/1367497.1367510

Potter, F., Dohrmann, K., Rockstroh, B., & Crombach, A. (2024). Refugees' integration and emotional distress over the course of 9 months. *Front. Psychol.*, *15*, 1459934. https://doi.org/10.3389/fpsyg.2024.1459934

Priemer, J., & Schmidt, M. (2019). Flüchtlingshilfe in der organisierten zivilgesellschaft. *Forum Wohnen und Stadtentwicklung*, *6*, 331–4. https://www.vhw.de/fileadmin/user_upload/08_publikationen/verbandszeitschrift/FWS/2019/6_2019/FWS_6_19_Priemer_Schmidt.pdf

Razai, M. S., Berry-Hart, T., Tawakoli, A. R. A., Ahmad, A., & Hargreaves, S. (2023). Silent suffering of Afghanistan's LGBT+ community. *Lancet Global Health*, *11*(9), e1331–e1332. https://doi.org/10.1016/S2214-109X(23)00316-9

Schneider, G., Segadlo, N., & Leue, M. (2020). Forty-eight shades of Germany: Positive and negative discrimination in federal asylum decision making. *German Politics*, *29*(4), 564–581. https://doi.org/10.1080/09644008.2019.1707810

Siegel, A., Wolff, J., & Weinstein, J. (2024). #asylum: How syrian refugees engage with online information. *Journal of Quantitative Description: Digital Media*, *4*. https://doi.org/10.51685/jqd.2024.013

Siegert, M., Tanis, K., Ette, A., & Sauer, L. (2023). Entwicklung der Wohnsituation ukrainischer Geflüchteter in Deutschland. *Kurzanalyse des BAMF-Forschungszentrums / BAMF-Kurzanalyse*. https://doi.org/10.48570/bamf.fz.ka.03/2023.d.11/2023.ukrwohnsituation.1.0

Stapf, T. (2019). *Migration / Digital: Die Bedeutung der Sozialen Medien für Ankommen, Orientierung und Teilhabe von Neuzugewanderten in Deutschland*. Mensch und Buch Verlag; Minor, Projektkontor für Bildung und Forschung. https://minor-kontor.de/wp-content/uploads/2020/09/NiB_Migration_Digital_Text_Web_20-09-29.pdf

StataCorp. (2023). Stata 18 base reference manual - margins. *Stata Press*. https://www.stata.com/manuals/cmmargins.pdf

Tjaden, J. (2021). Measuring migration 2.0: A review of digital data sources. *Comparative Migration Studies*, *9*(1). https://doi.org/10.1186/s40878-021-00273-x

Tjaden, J., Haarmann, E., & Savaskan, N. (2022). Experimental evidence on improving covid-19 vaccine outreach among migrant communities on social media. *Scientific reports*, *12*(1), 16256. https://doi.org/10.1038/s41598-022-20340-2

Tjaden, J., & Spörlein, C. (2023). How much do “local policies” matter for refugee integration? an analytical model and evidence from a highly decentralized country. *International Migration Review*, 01979183231205561.

Tür an Tür - Digitalfabrik gGmbH. (2023). Wirkungsbericht 2022. https://integreat-app.de/wp-content/uploads/2023/05/Digitalfabrik-Wirkungsbericht-2022.pdf

Tür an Tür - Digitalfabrik gGmbH. (2024). Jährliche Umfrage | Integreat - die digitale Integrationsplattform 2024 [Unpublished material. Access upon request by Tür an Tür - Digitalfabrik gGmbH].

Supplemental materials

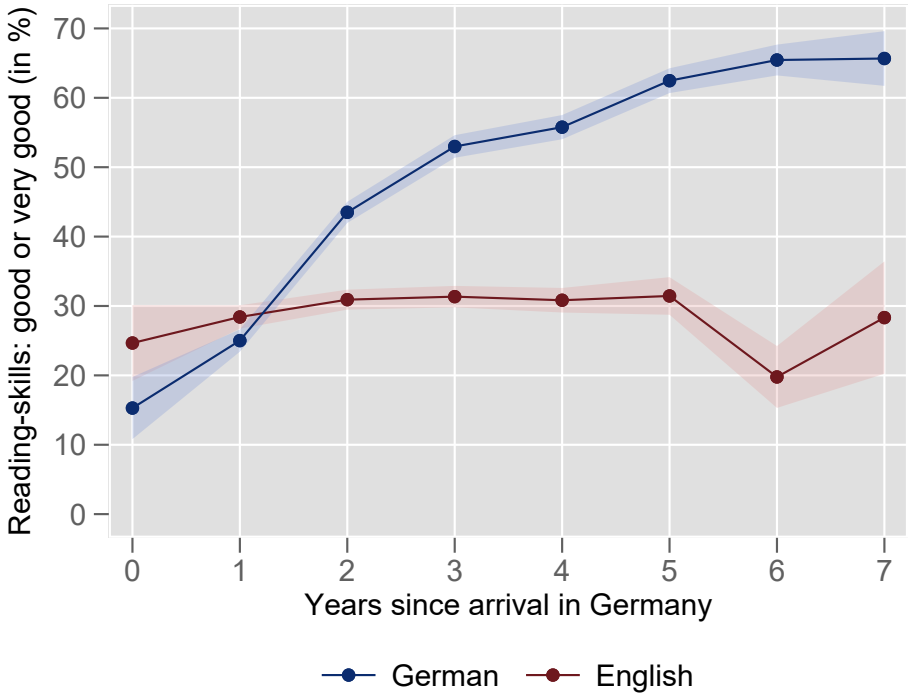

Figure A1. German & English reading skills (2014-2021): Weighted share of refugee respondents stating good or very good reading skills (in %) with 95% confidence intervals among the literate sub-population with (very) good reading skills reported in the first language. *Source*: SOEP v38.1 2016-2021 (Goebel et al., 2023), own calculations.

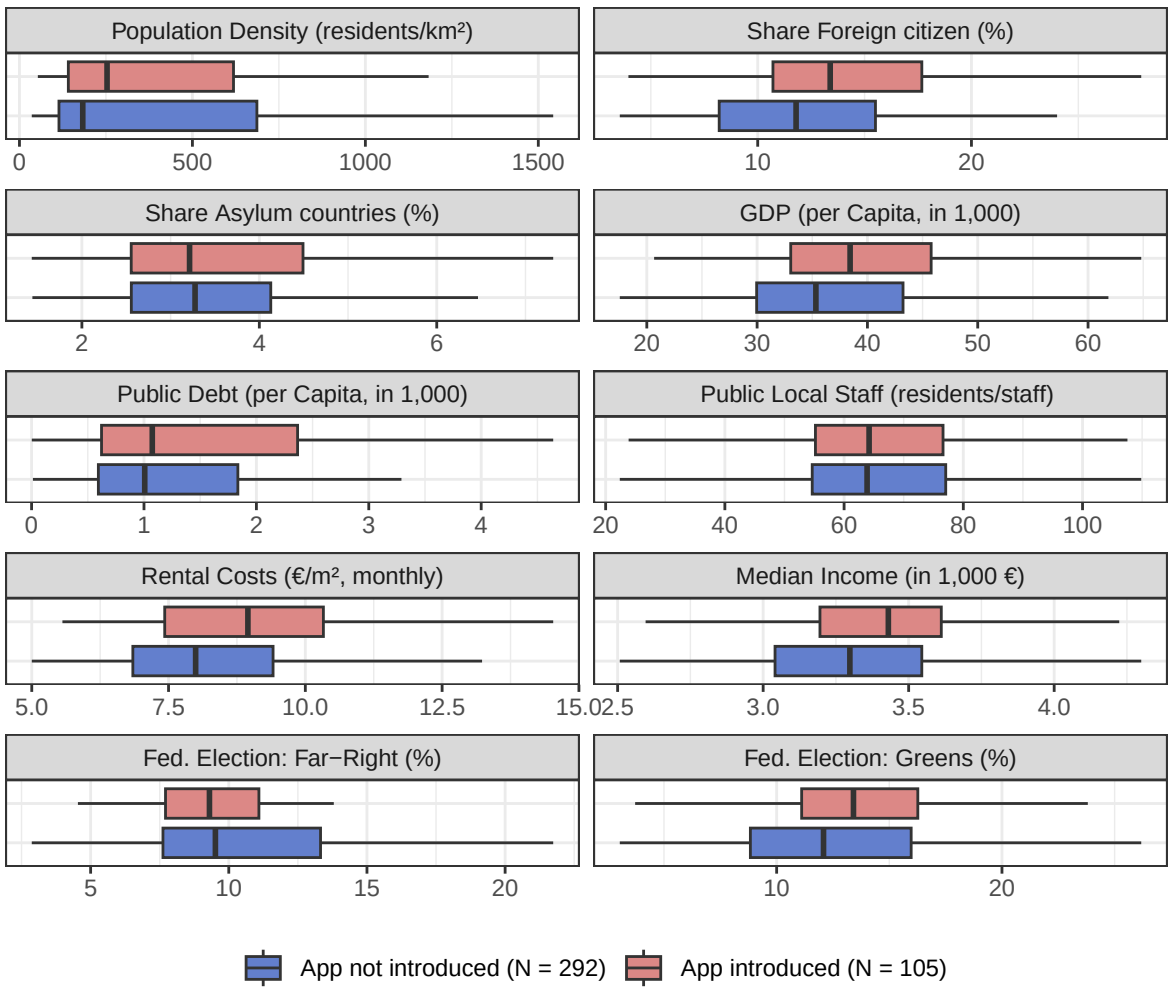

Figure A2. Characteristics of (non-)participating regions: Continuous county-level characteristics (N = 397), summarized as box plots (min, Q25, median, Q75, max), by participation status (app ever introduced by May 2024). *Source*: Tür an Tür - Digitalfabrik gGmbH (2023); Bundesinstitut für Bau-, Stadt- und Raumforschung (BBSR) (2024). Own calculation and illustration.

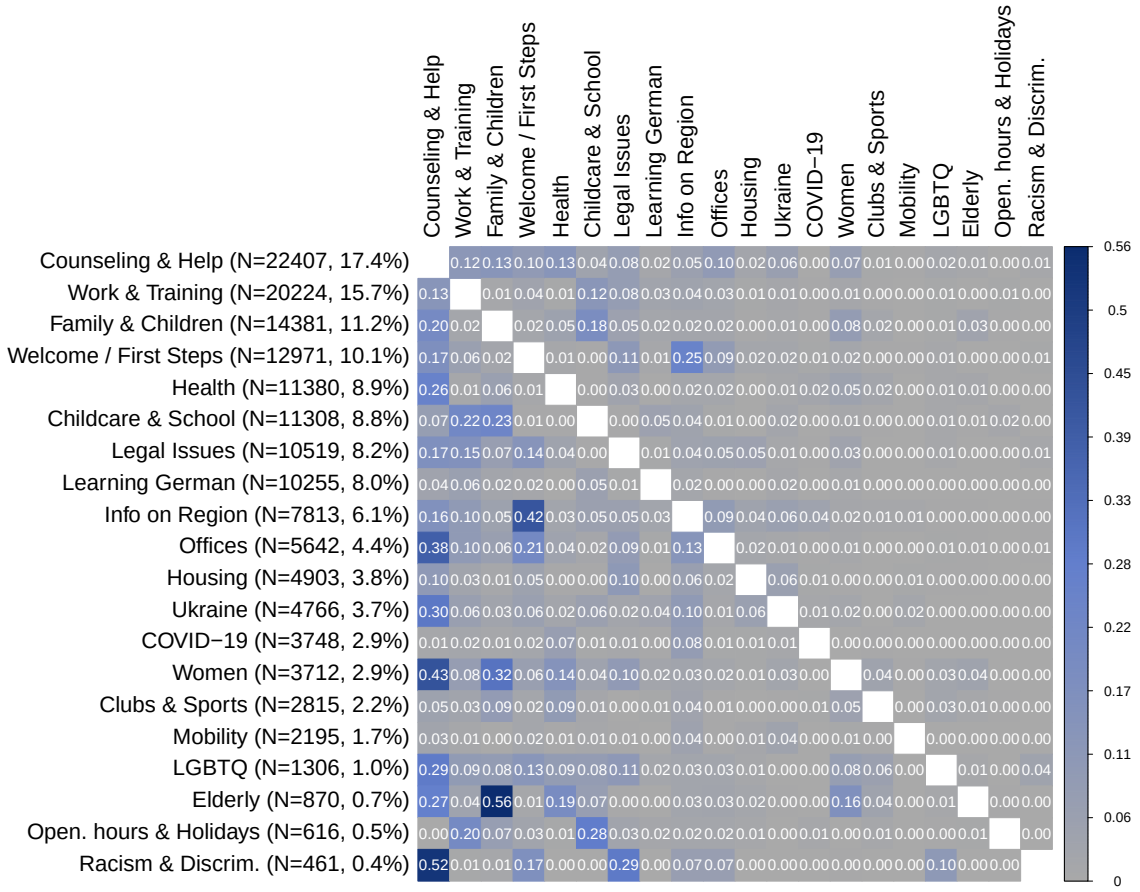

Figure A3. Coincidence matrix of topics (2021-2024): Brackets report the number of unique URL strings containing the respective topic keywords, and the percentage relative to all unique URLs. Cell values indicate the share (0-1) of co-occurrence between categories, given that categories in rows are observed. **Reading example**: Topic *Counseling* is identified in 23,768 URLs, representing 17.4% of URLs overall. If counseling is observed, it shares 12% of the cases that are also identified as belonging to the topic *Work & Training*. **Source**: Tür an Tür - Digitalfabrik gGmbH (2023), own calculations.

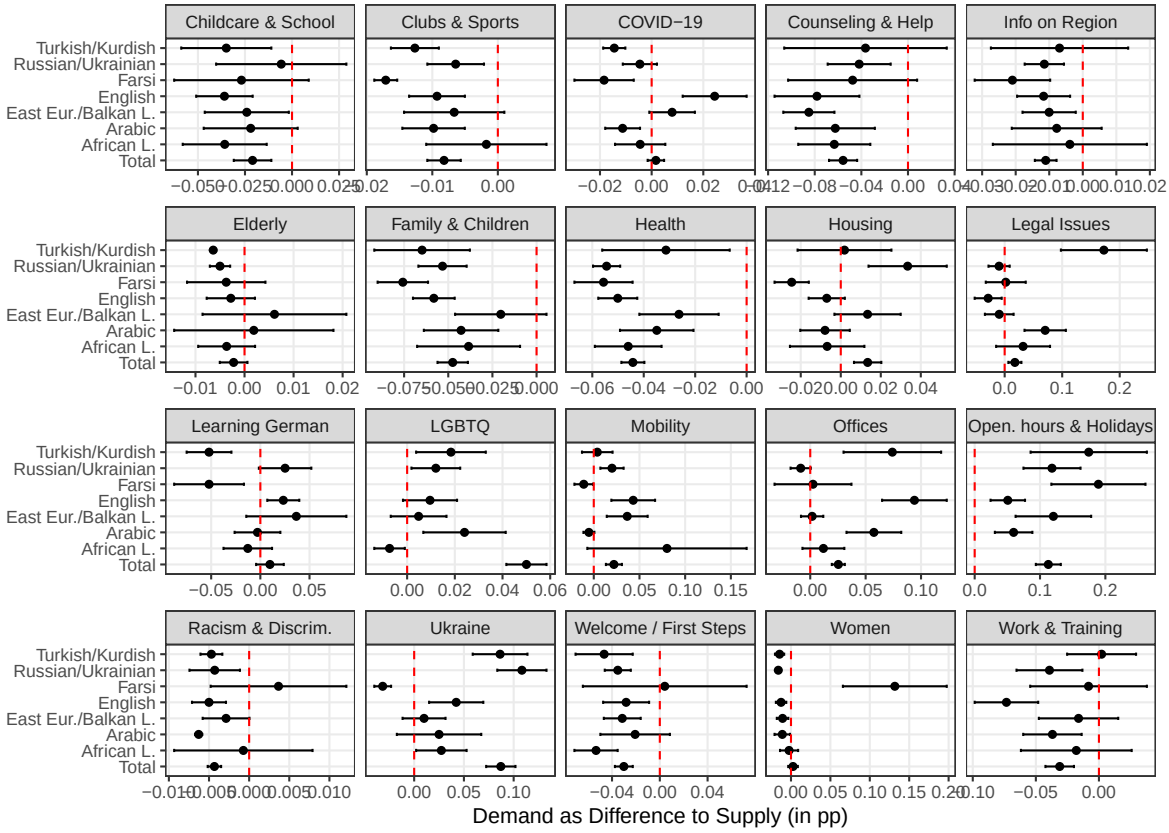

Figure A4. Prediction of page visits relative to supply: Percentage-point differences (probability of page visits minus median supply) by used language; vertical dashed line at 0 denotes perfect alignment.

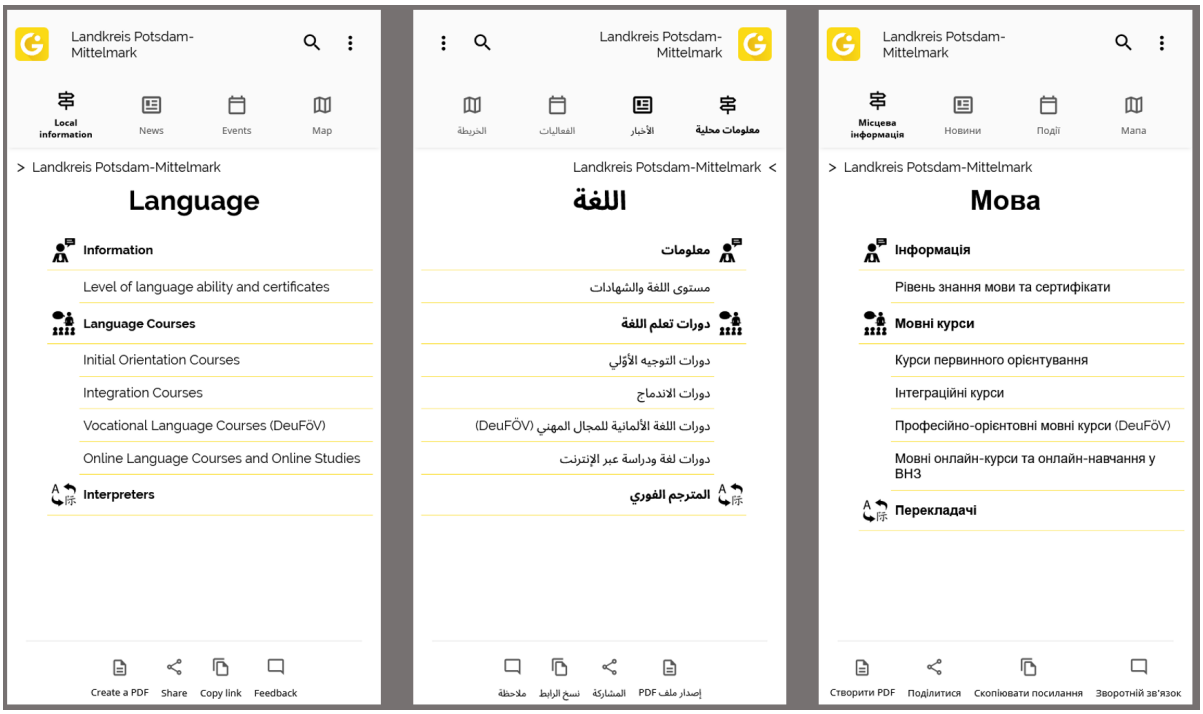

Figure A5. First level of information on language: First level of information – after selection of 'Language' tile on home-screen – of three exemplary platform versions (region-language combinations).

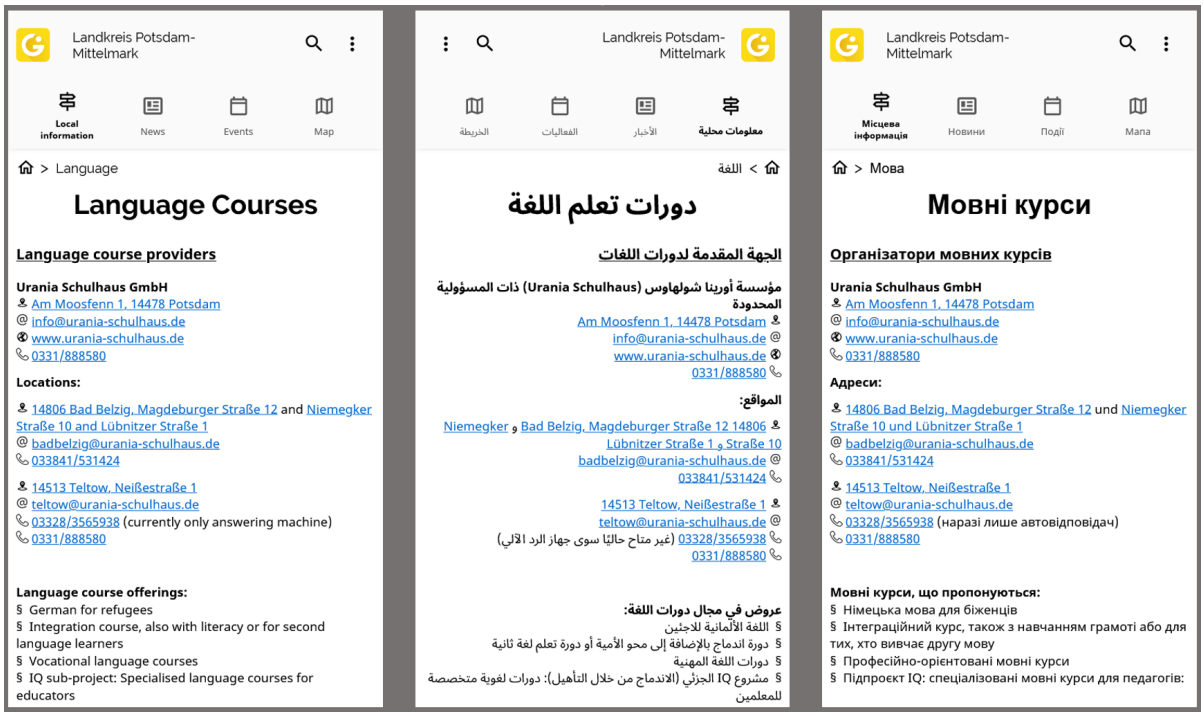

Figure A6. Second level of information on language: Language courses.: Second level of information – after selection of 'Language Courses' on previous screen – of three exemplary platform versions (region-language combinations).

Table A1: Topic identification.

Topic	Search Query (RegEx)
Learning German	\blearn german\b|\blanguage\b|integration(?:course|skurs)| (?:language|sprach)(?:\s|-)?(?:course|kurs)| \bsprache\b
Health	health|medicine|\bdoctor\b|hospital|\barzt\b|physician
Counseling & Help	counsel(?:ing|ling)|advice|help|assistance|consultation|consultancy
Work & Training	work|\bjob\b|employment|apprenticeship|vocation
Mobility	\btrain\b|transport|\bbus\b|\bcar\b|mobility
Childcare & School	child(?:\s|-)?care|kindergarten|day(?:\s|-)?care|\bkita\b|school
Legal Issues	asylum|legal|rights|\blaw\b|\blaws\b|rights
Housing	housing|apartment|\bflat\b|accommodation
Offices	\boffice\b|authorit(?:y|ies)
Welcome / First Steps	welcome|first steps
Ukraine	ukraine
COVID-19	corona|covid
Clubs & Sports	club|sports|football|soccer|pools|swimming|tennis
LGBTQ	lgbt|trans(?!port|lat)|\bgay\b|lesbian|bisexual|bi-sexual|queer
Women	wom[ae]n
Family & Children	children|family
Racism & Discrim.	racism|racist|discrim
Elderly	elder|senior
Info on Region	kreis|stadt|county|district|city
Open. hours & Holidays	opening hour(?:s)|holiday(?:s)|vacation

Table A2: Topics-Top 20 occurring words in URL.

Topic (# URLs)	Top 20-words (% observed in topic)
Counseling & Help (N = 22407)	advice (39.3%), help (29.9%), assistance (24.4%), counseling (21.8%), centers (12.4%), health (12.3%), support (11.3%), family (10.9%), information (9.9%), consultations (9.3%), welcome (9.0%), children (8.5%), work (7.8%), youth (7.5%), people (7.2%), counselling (7.2%), services (7.0%), authorities (6.9%), migration (6.8%), centres (6.7%)
Work & Training (N = 20224)	work (75.9%), training (40.5%), education (36.6%), vocational (21.8%), job (14.7%), information (13.7%), study (13.5%), employment (12.4%), school (10.6%), professional (9.7%), studies (8.2%), higher (6.4%), advice (5.9%), university (5.7%), general (5.5%), center (5.2%), recognition (4.9%), career (4.9%), find (4.7%), important (4.6%)
Family & Children (N = 14381)	family (79.8%), children (67.8%), youth (36.9%), young (13.4%), families (12.7%), pregnancy (11.4%), people (10.7%), support (10.1%), care (8.3%), school (8.2%), offers (7.9%), women (7.7%), advice (7.0%), child (6.7%), counseling (6.4%), childcare (6.0%), help (5.9%), birth (5.9%), education (5.8%), leisure (5.5%)
Welcome / First Steps (N = 12971)	welcome (91.0%), district (17.8%), important (16.2%), germany (12.6%), first (11.8%), steps (10.8%), offices (8.0%), advice (7.9%), office (6.9%), asylum (6.7%), assistance (6.5%), information (6.4%), living (6.3%), help (4.4%), city (4.3%), together (4.2%), social (4.2%), migration (3.9%), integration (3.5%), area (3.5%)

Table A2: Topics-Top 20 occurring words in URL.

Topic (# URLs)	Top 20-words (% observed in topic)
Health (N = 11380)	health (95.4%), information (13.0%), general (9.0%), help (8.5%), care (8.3%), counseling (7.9%), assistance (7.9%), insurance (6.7%), emergency (6.6%), centers (6.5%), advice (6.1%), pregnancy (5.3%), mental (5.2%), people (5.0%), services (5.0%), medical (4.9%), doctor (4.9%), system (4.8%), offers (4.6%), pharmacies (4.1%)
Childcare & School (N = 11308)	school (79.4%), education (49.2%), schools (20.7%), children (17.7%), training (16.1%), family (15.6%), study (14.6%), work (14.1%), kindergarten (12.9%), childcare (12.9%), care (12.1%), vocational (10.9%), child (10.3%), information (9.8%), studies (9.3%), 6 (8.5%), general (8.5%), youth (8.0%), years (7.9%), day (6.3%)
Legal Issues (N = 10519)	asylum (57.7%), legal (31.5%), information (18.7%), rights (18.5%), refugees (13.4%), welcome (12.4%), residence (12.3%), work (11.8%), germany (9.7%), law (9.2%), procedure (9.1%), advice (7.9%), seekers (7.7%), migration (6.9%), issues (6.5%), immigration (5.8%), assistance (5.6%), training (5.6%), important (5.4%), application (5.4%)
Learning German (N = 10255)	language (92.9%), german (51.1%), courses (42.3%), learn (19.5%), learning (12.2%), integration (12.2%), course (9.5%), education (8.9%), online (7.1%), offers (5.7%), information (5.6%), level (4.4%), orientation (4.0%), school (3.9%), study (3.5%), classes (3.4%), initial (3.1%), certificates (2.9%), work (2.9%), training (2.8%)

Table A2: Topics-Top 20 occurring words in URL.

Topic (# URLs)	Top 20-words (% observed in topic)
Info on Region (N = 7813)	district (61.6%), welcome (40.5%), city (25.2%), information (9.9%), important (9.0%), cities (7.4%), advice (7.3%), office (7.1%), life (6.4%), work (6.0%), ukraine (6.0%), education (5.0%), living (4.9%), help (4.9%), integration (4.8%), munich (4.8%), county (4.5%), aschaffenburg (4.5%), hof (4.5%), municipalities (4.5%)
Offices (N = 5642)	office (55.4%), authorities (50.0%), important (32.8%), offices (26.9%), advice (20.9%), welcome (18.3%), youth (10.8%), social (10.2%), district (9.7%), welfare (9.6%), immigration (9.6%), migration (8.2%), authority (7.9%), services (7.8%), foreigners (6.8%), integration (6.4%), refugees (6.4%), counseling (6.0%), asylum (5.9%), counselling (5.9%)
Housing (N = 4903)	housing (49.2%), apartment (30.2%), life (28.0%), accommodation (24.8%), everyday (18.3%), living (13.3%), search (12.5%), asylum (9.5%), reside (9.2%), residence (8.9%), information (7.8%), rent (7.4%), daily (7.3%), social (6.9%), ukraine (5.7%), own (5.6%), leisure (5.3%), live (4.8%), find (4.3%), moving (4.1%), refugees (4.1%)
Ukraine (N = 4766)	ukraine (98.3%), information (36.8%), help (25.7%), people (15.2%), crisis (12.2%), refugees (11.9%), war (10.0%), ukrainian (9.0%), current (8.1%), everyday (6.9%), district (6.9%), life (6.7%), aid (6.3%), arrival (5.3%), school (4.8%), work (4.5%), accommodation (4.1%), topics (4.1%), first (3.6%), augsburg (3.6%)

Table A2: Topics-Top 20 occurring words in URL.

Topic (# URLs)	Top 20-words (% observed in topic)
COVID-19 (N = 3748)	corona (71.3%), virus (63.1%), coronavirus (35.6%), information (32.0%), vaccination (19.3%), against (9.7%), covid (8.0%), app (7.5%), 19 (7.0%), district (7.0%), health (6.5%), current (6.2%), warn (5.7%), rules (3.5%), vaccinate (2.9%), test (2.8%), important (2.5%), crown (2.4%), misinformation (2.2%), general (2.1%)
Women (N = 3712)	women (85.2%), children (25.8%), girls (23.1%), family (22.4%), offers (21.8%), help (17.6%), women's (15.5%), advice (14.8%), violence (13.6%), health (13.2%), youth (13.1%), support (11.3%), counseling (10.2%), assistance (9.6%), families (9.1%), rights (8.5%), young (8.2%), services (8.0%), culture (7.8%), leisure (7.6%)
Clubs & Sports (N = 2815)	sports (71.9%), leisure (52.8%), culture (44.9%), clubs (24.6%), time (22.6%), life (16.4%), free (16.0%), sport (13.7%), everyday (13.0%), offers (9.9%), health (9.4%), swimming (8.5%), pools (8.1%), children (6.9%), people (6.7%), activities (6.6%), youth (6.2%), meeting (5.3%), freetime (4.8%), associations (4.5%)
Mobility (N = 2195)	mobility (76.9%), life (62.9%), everyday (52.1%), transport (21.2%), daily (17.5%), public (16.4%), car (12.3%), bus (11.1%), bicycle (10.3%), train (9.6%), license (9.4%), driving (9.4%), germany (8.9%), transportation (7.5%), living (7.4%), leisure (6.3%), means (5.5%), time (5.1%), travel (5.0%), free (4.3%)

Table A2: Topics-Top 20 occurring words in URL.

Topic (# URLs)	Top 20-words (% observed in topic)
LGBTQ (N = 1306)	bisexual (22.8%), lgbti (22.4%), gay (21.7%), lesbian (19.8%), life (19.2%), leisure (16.5%), trans (16.0%), culture (14.3%), advice (12.6%), welcome (12.3%), everyday (11.9%), offers (11.8%), inter (11.6%), queer (11.3%), lgbt (10.7%), group (10.2%), lesbians (9.9%), people (9.5%), sport (9.2%), germany (8.9%), transition (8.9%)
Elderly (N = 870)	family (47.9%), elderly (47.5%), children (37.2%), seniors (36.3%), care (24.8%), youth (22.6%), people (22.3%), offers (19.9%), health (18.4%), senior (15.5%), young (14.9%), women (13.0%), families (12.5%), information (10.8%), advice (10.5%), citizens (10.1%), counseling (7.6%), centers (7.0%), social (6.1%), leisure (5.9%)
Open. hours & Holidays (N = 616)	holidays (88.3%), life (42.5%), hours (36.2%), everyday (34.7%), public (30.2%), school (28.2%), opening (23.1%), time (13.3%), working (13.3%), education (11.4%), daily (11.0%), information (9.3%), germany (8.8%), bavaria (8.1%), news (7.8%), leisure (7.6%), work (6.5%), vacation (6.3%), children (5.8%), latest (5.5%)
Racism & Discrim. (N = 461)	discrimination (86.3%), anti (42.3%), advice (29.1%), rights (27.5%), assistance (22.6%), support (22.1%), violence (18.7%), legal (16.7%), inclusion (16.5%), welcome (16.5%), right (16.3%), racism (16.1%), against (15.8%), help (13.9%), wing (12.4%), counseling (11.9%), life (9.8%), information (9.3%), germany (7.8%), victims (7.6%)

Table A3: Reduction of dataset.

Step	Explanation	OBS	Excluded (N)	Excluded (%)	Excluded (Cum.%)
1	Raw count.	11813123			0.0
2	Remove duplicate rows, defined as identical URL and time (by seconds).	11442666	370457	3.1	3.1
3	Remove rows if URL erroneous, defined by non-valid language-code.	11439196	3470	0.0	3.2
4	Remove rows for very rare languages <=100 observations.	11438980	216	0.0	3.2
5	Remove rows if base-URL represents test-environments or other non-relevant sites.	10751030	687950	5.8	9.0
6	Remove rows if city sub-page does not show up on sundays: indicates test-use only.	10721371	29659	0.3	9.2
7	Keep only city sub-pages with at least 1,500 observations.	10710585	10786	0.1	9.3
8	Remove rows for city sub-pages indicating regions outside Germany, non-counties or test-environments.	10666671	43914	0.4	9.7
9	Remove rows that only represent start-page visits, i.e. no topic-specific content.	9637789	1028882	8.7	18.4
10	Remove rows for city sub-pages with >90% Offline usage. Online research suggests that they never went live.	9593831	43958	0.4	18.8
11	Remove rows that fall into lead or trail periods within city- language groups with <10 observations or with <10 % of expected visits on weekends, assuming univariate distribution across weekdays.	9441368	152463	1.3	20.1
12	Remove rows that fall into city-language series shorter than three months.	9409592	31776	0.3	20.3

Table A3: Reduction of dataset.

Step	Explanation	OBS	Excluded (N)	Excluded (%)	Excluded (Cum.%)
13	URL contains no information on topics: Offline-usage & all visits prior 08/2021	6955500	2454092	20.8	41.1
14	Delete further languages: German (N = 3,198,832), Centr/South Europe (N = 102,648), Brazilian Portuguese (N = 253) and Chinese (N = 277).	3653490	3302010	28.0	69.1
15	Delete cases if one topic represents >80% of all URLs.	3571680	81810	0.7	69.8
16	Delete cases if prior 09/2021 or after 02/2024 as Topic Supply-Identifaction works less well on Edges.	2959802	611878	5.2	74.9

Summary statistics (RQ2, RQ3)

Table A4 presents summary statistics for all logistic models as outlined in Equation 1. The models vary between 885,103 observations (*Racism & Discrimination*) and 2,928,689 observations (*Work & Training*), which is a result of the design restriction to include cases only if a respective topic page could have been visited, i.e., $TopicSupply_{r,l,t} > 0$.

This is also reflected in the number of platform versions, considered per model, which varies between 165 and 634, all of them still providing a sufficient amount of variation to compare levels of supply. The median of $TopicSupply_{r,l,t}$ across i varies by model from 0.7 to 15.8%, similar to that observed on the regional level.[16] The analysis years are fairly similar across topics, with the exception of *COVID-19* which is less often observed in later years. The median number of distinct URLs for platform versions observed in i varies from 114 to 158 and appears to be higher for topics that are less frequently provided, like *The Elderly* or *Racism & Discrimination*. This indicates that these topics tend to be represented by regions with a generally larger overall information supply.

Column explanation:

- Total (N) presents the number of observations per model.
- Visit (%) presents the share of observations that is identified as the topic in question.
- Year (Mean) returns the mean year across all observations.
- #URLs (p50) represents a median indicator for the app size across all observations. Specifically, it measures the median of the URLs available to i, across all observations.
- %Topic (p50) returns the median of $TopicSupply_{r,l,t}$ in i, across all observations.
- Region-Lang (Dist.) returns the number of unique platform versions, i.e., region-language combinations, observed in the respective model.

[16] The median values are not identical to Figure 4 because the platform versions may vary in individual page visit size, i.e., platforms with more page visits contribute more strongly here when directly calculated from i as compared to Figure 4 where aggregated on the level of platform-version-days r, l, t for emphasizing a cross-regional comparison and administrations decisions.

Table A4: Summary statistics for each regression model

DV	Total (N)	Visit (%)	Year (Mean)	#URLs (p50)	%Topic (p50)	Region-Lang (Dist.)
Overall	2959802		2022.631	114		644
Work & Training	2928689	11.761	2022.636	115	13.485	633
Counseling & Help	2918955	12.043	2022.636	115	15.789	634
Family & Children	2874051	7.418	2022.635	117	11.741	614
Learning German	2873500	11.377	2022.643	117	8.159	621
Childcare & School	2850456	7.420	2022.643	118	8.589	608
Health	2818202	4.697	2022.632	119	8.421	609
Welcome / First Steps	2809700	10.368	2022.638	118	9.091	613
Legal Issues	2710351	12.242	2022.653	120	7.764	586
Info on Region	2706629	3.469	2022.640	121	3.374	568
Housing	2640500	8.053	2022.647	123	3.774	529
Offices	2402629	6.057	2022.648	127	2.508	526
Mobility	2402532	5.169	2022.640	130	2.222	483
Women	2228264	2.770	2022.655	132	2.128	464
Clubs & Sports	2094988	1.716	2022.600	132	2.048	383
LGBTQ	1773393	6.660	2022.742	136	1.395	334
COVID-19	1637852	3.813	2022.196	124	2.756	463
Open. hours & Holidays	1634083	12.372	2022.635	146	0.820	289
Ukraine	1343875	17.999	2022.667	129	4.054	337
Elderly	929562	1.502	2022.708	156	0.694	187
Racism & Discrim.	885103	0.321	2022.680	158	0.667	165

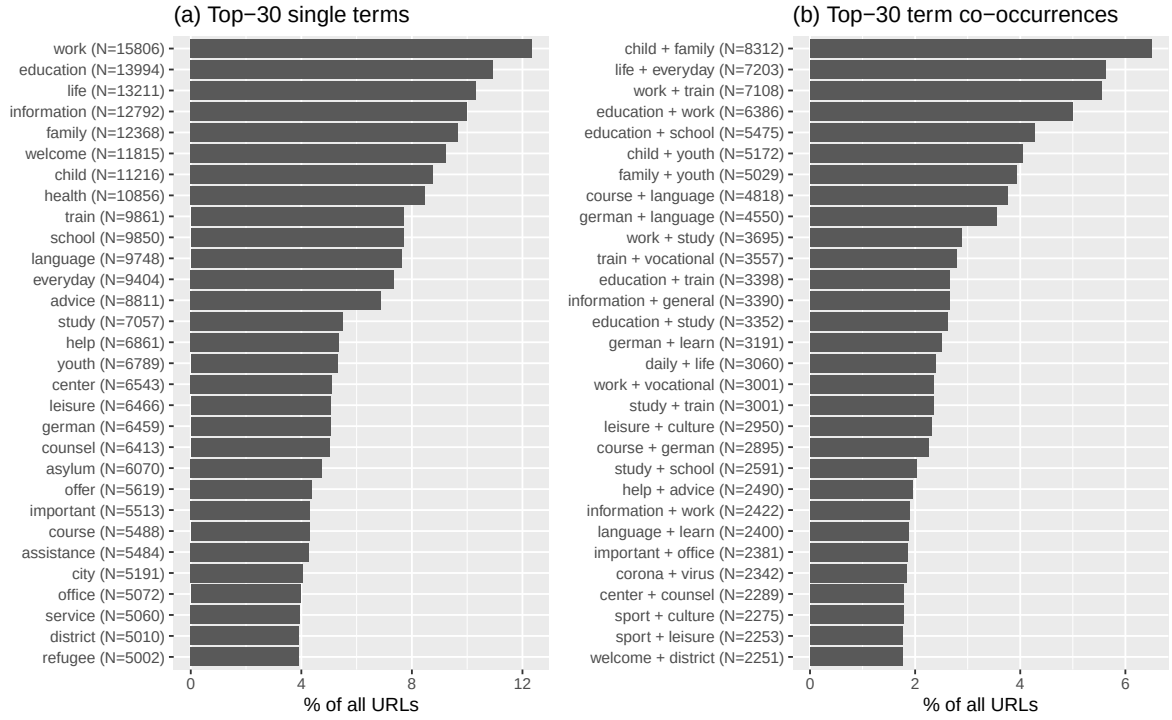

Figure A7. URL-level prevalence of terms: Displayed are URL-level (N = 128,420) statistics on (a) the prevalence of single terms and (b) co-occurrences of two terms in any order. Restricted to URL parts containing page contents. Further cleaning: filling words removed (snowball stop word list), lemmatization of words ('textstem' R-package), removal of duplicate terms within one URL, removal of city names.

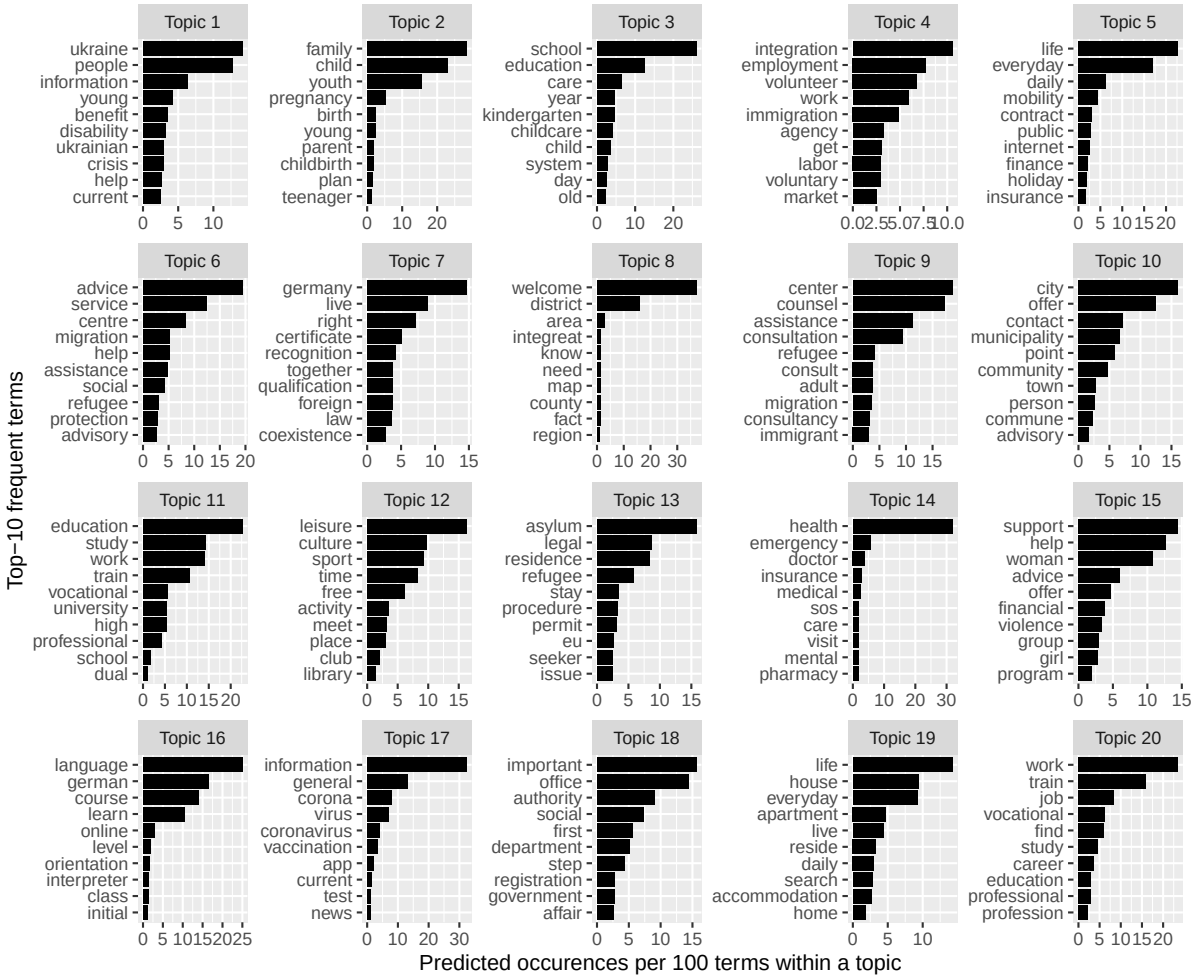

Figure A8. Unsupervised topic modeling (LDA): Most frequent terms per topic.: Unsupervised topic modeling (Latent Dirichlet Allocation, Gibbs sampler), applied to a document-term matrix built from terms contained in URLs where each document equals one URL. For each topic, the ten terms with the highest topic-word probability (β) are shown. Bar length is $\beta \times 100$ and can therefore be read as the predicted number of occurrences per 100 terms within that topic. Further cleaning: filling words removed (snowball stop word list), lemmatization of words ('textstem' R-package), removal of duplicate terms within one URL, removal of city names.

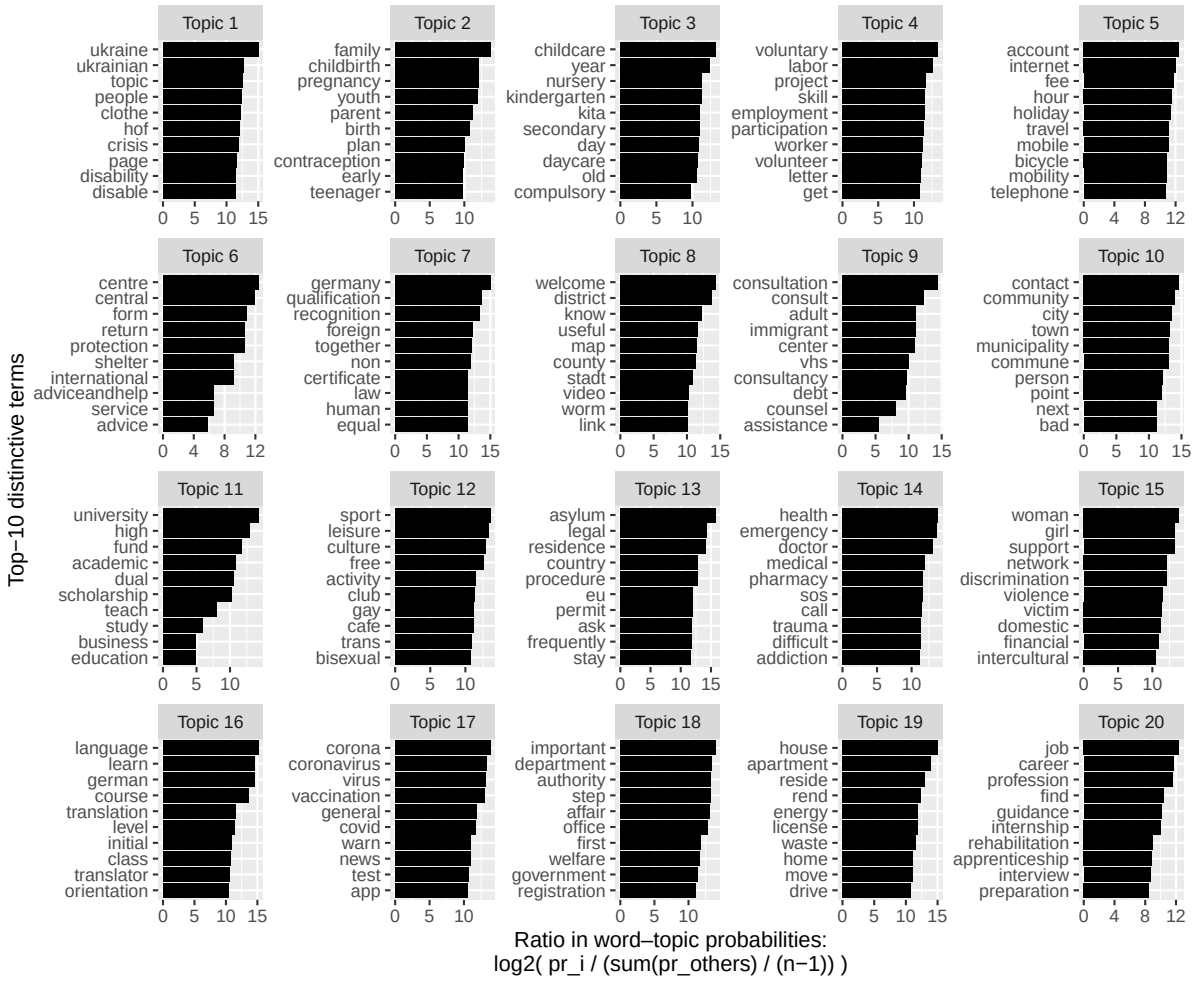

Figure A9. Unsupervised topic modeling (LDA): Most distinctive terms: Unsupervised topic modeling (Latent Dirichlet Allocation, Gibbs sampler), applied to a document-term matrix built from terms contained in URLs where each document equals one URL. Panels show the 10 most distinctive terms per topic, ranked by $log2[\beta topic/mean(\beta othertopics)]$; a value L means the term is 2^L times more likely in that topic. For example, the word 'ukraine' in topic 1 has a log-ratio of 15, meaning that its probability within topic 1 is about $2^{15} \approx 33,000$ times higher than its average probability across all the other topics. Rare terms with $\beta \leq 0.005$ were omitted here to reduce noise.

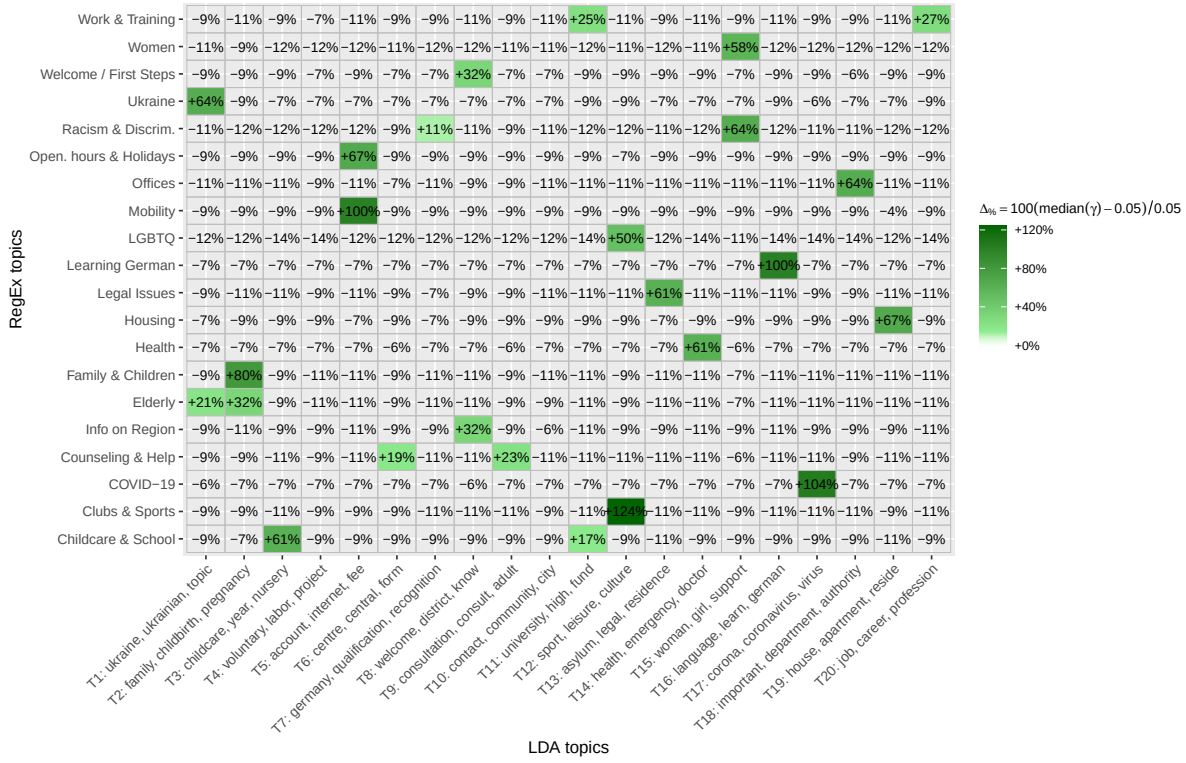

Figure A10. Supervised (RegEx) versus unsupervised (LDA) topic modeling: Each cell reports the percent deviation of the median URL-topic probability γ among URLs tagged by a RegEx topic – from the *'no-signal'* baseline of 1/K = 0.05 (K = 20): $\Delta\% = 100 \times (\text{median}(\gamma) - 0.05)/0.05$. **Example**: For "Clubs & Sports," the **median** γ for the LDA topic labeled "sport, leisure, culture" is 0.10, so $\Delta\% = 100 \times (0.10 - 0.05)/0.05 = +100\%$ (twice the *'no-signal'* baseline).

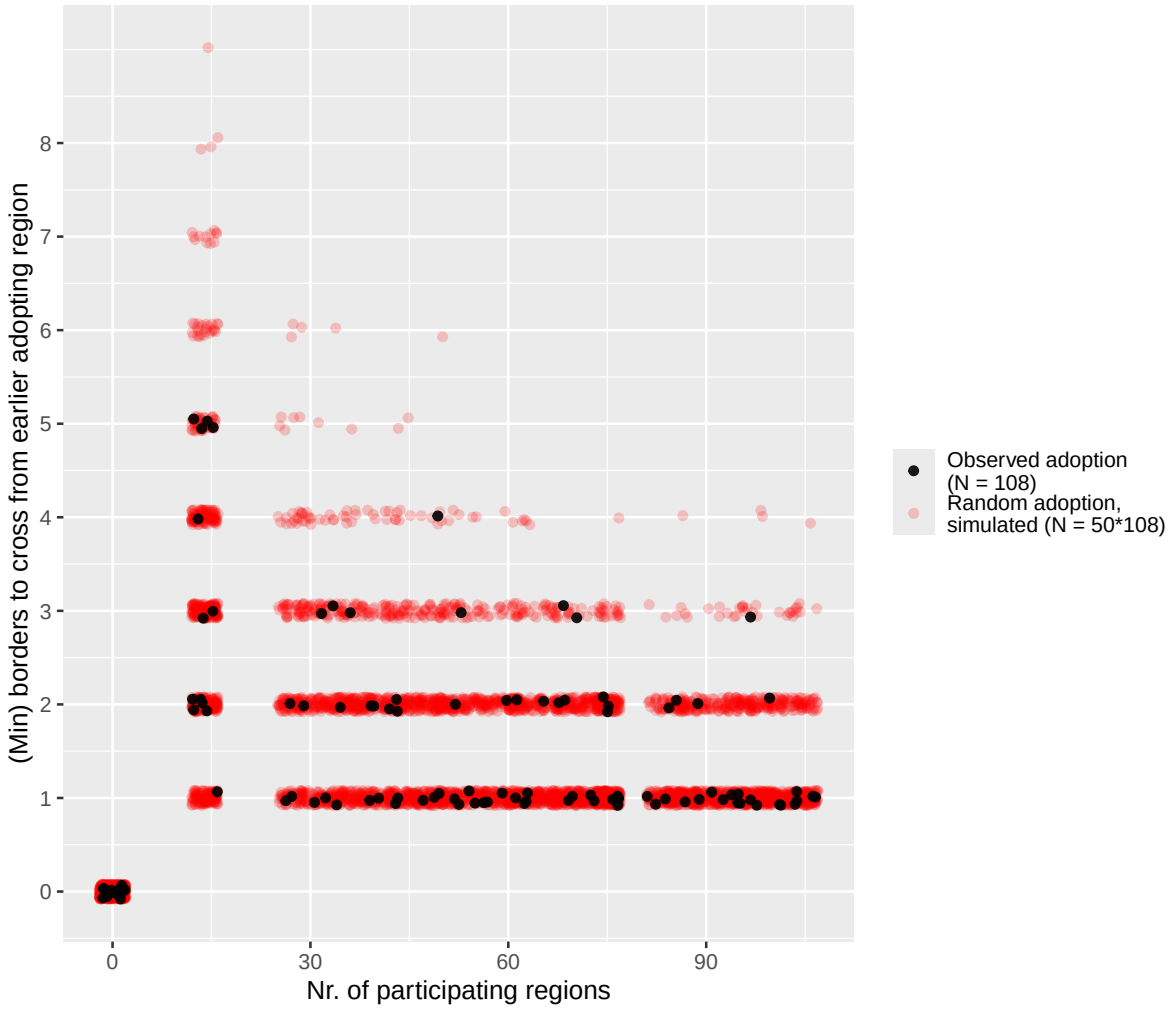

Figure A11. Observed and simulated diffusion of app-adoption across German districts: Each point is one district in the month it first adopts. The x-axis shows the cumulative number of adopters up to and including that month; the y-axis is the shortest path border distance: the minimum number of district borders that must be crossed from the adopting district to reach any earlier adopting district. Distances are calculated based on NUTS-3 county shape files for Germany and their adjacencies. First adopters are set to 0 by definition. Black points are the observed adoptions (N = 108 counties). Red points are simulated adoptions (N = 108*50 counties) that preserve the observed monthly counts of new adopters per simulation while selecting adopters uniformly at random from the pool of not-yet-adopting districts.

Table A5: Difference in walk distance between observed and simulated adoptions

term	Coef.	P_value	N_observed	N_simulated
Observed vs. Simulated	-0.033	0.663	94	4700

Note. Coefficient reports the average difference in walk distance (observed minus simulated) with treated-cumulative fixed effects. Standard errors are robust. A negative coefficient indicates shorter observed diffusion paths relative to random allocation.