Directional Semantic Drift Across Grokipedia Versions: A Diffusion-Manifold Analysis

Veronika Batzdorfer
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
veronika.batzdorfer@kit.edu

Oliver Tölkes
Independent Researcher
Cologne, Germany
info@toelkes.org

Abstract

Large language models increasingly mediate access to factual knowledge, yet little is known about how their representations diverge from established encyclopedic sources. We quantify semantic drift between Wikipedia and Grokipedia–a large language model– generated encyclopedia explicitly designed to counter perceived Wikipedia biases–as well as across two Grokipedia versions released before and after a platform update (v0.1, v0.2). Focusing on articles about cities in the United States and Germany, we embed 1,387 matched entries, project them onto a shared diffusion manifold, and operationalize semantic drift as Euclidean displacement in latent space. Both cross-sectional and longitudinal analyses reveal systematic structure in this drift: larger cities and articles containing politically salient framing exhibit significantly greater divergence. Regression models further show that changes in article length and political content predict drift even after controlling for population size, country, and time. Articles in Grokipedia v0.1 that were explicitly attributed as adapted from Wikipedia lost this attribution in v0.2 and exhibited a pronounced and statistically significant semantic shift, indicating a new article generation. The proposed diffusion-manifold framework provides a scalable testbed for researching semantic bias and temporal drift in AI-mediated knowledge.

CCS Concepts

• **Mathematics of computing** → *Information theory*; • **Information systems** → **World Wide Web**; **Information retrieval**; • **Applied computing** → **Psychology**; **Document management and text processing**.

Keywords

Wikipedia, Grokipedia, v0.2, v0.1, Generative AI, large language models, information retrieval, diffusion manifold, semantic drift, temporal analyses

ACM Reference Format:
Veronika Batzdorfer and Oliver Tölkes. 2026. Directional Semantic Drift Across Grokipedia Versions: A Diffusion-Manifold Analysis. In *18th ACM Web Science Conference (WebSci '26), May 26–29, 2026, Braunschweig, Germany*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3795766.3799761

WebSci '26, Braunschweig, Germany

ACM ISBN 979-8-4007-2504-3/26/05
https://doi.org/10.1145/3795766.3799761

1 Introduction

Large language models (LLMs) are increasingly embedded in systems of public knowledge access, ranging from web search to encyclopedic reference services [9]. These systems do not merely retrieve information but actively mediate, summarize, and reframe knowledge, raising foundational questions about how collective representations of the world are produced and transformed [4, 9, 12]. In the current landscape of AI-mediated knowledge, Wikipedia holds an exceptional position. The Wikimedia Foundation reports that its content is routinely used in the training of large language models and often constitutes a dominant component of their training corpora [3, 6]. Yet, at the same time, Wikipedia remains governed by editorial norms—such as verifiability, and attribution—that differ fundamentally from the statistical and generative mechanisms of LLMs [6]. While community-curated platforms such as Wikipedia have long served as a benchmark for encyclopedic knowledge, recent AI-generated platforms—such as Grokipedia, an LLM-driven encyclopedic service—introduce a fundamentally different mode of knowledge production based on generative inference rather than human editorial consensus. Existing research on Grokipedia-mediated knowledge has largely focused on macro-level corpus properties [10], topical or political bias [10, 14] in specific domains, or shifts in epistemological justifications [7]. By contrast, systematic comparisons of how LLMs represent specific knowledge domains relative to established encyclopedic sources remain limited.

We address this gap by comparing encyclopedic representations of prominent cities in Germany and the United States across Wikipedia and Grokipedia. Cities constitute a theoretically important and methodologically tractable domain: they are widely covered, unevenly documented, and often presumed to be relatively apolitical, while nevertheless encompassing contested topics such as migration, housing, and social policy. If generative and community sources diverge systematically in this apolitical knowledge domain, we could uncover a foundational bias in algorithmic knowledge systems. We treat generative encyclopedic systems as maintaining a dynamic internal representation shaped by training data, retrieval mechanisms, and platform updates.

We test two competing predictions about how AI-generated content diverges from human-curated sources, derived from research on cognitive bias and domain adaptation [5, 8, 11]. **Availability bias** predicts that abundant information amplifies rather than constrains distortion. When source material is extensive, generative systems have more "room" to selectively emphasize, reframe, or elaborate—producing outputs that drift further from any single reference point. Applied to encyclopedic content, this yields **H1: Cities with**

longer Wikipedia articles will show greater semantic divergence from Grokipedia (not less, as one might expect if more information improved accuracy). **Transfer bias** predicts that domain adaptation improves with better source integration. When a system gains access to higher-quality retrieval or updated modeling, its outputs should shift toward established reference standards—especially where previous coverage was poor and relied on weak priors. The November 21, 2025 platform update introduced a new model version and modified retrieval pipelines, providing a natural experiment. This yields **H2: Post-update Grokipedia summaries will move closer to Wikipedia, with the largest corrections for cities that previously had the sparsest coverage.** Both hypotheses are tested by measuring semantic displacement as Euclidean distance in a shared diffusion-manifold embedding of 1,387 matched city articles from Germany and the United States (Grokipedia v0.1 and v0.2, with corresponding Wikipedia entries). The data and code are made available for reproducibility[1].

2 Related works

Newly emerging platforms such as Grokipedia now generate encyclopedic content at scale, yet empirical comparisons with human-curated baselines remain sparse. Yasseri and Mohammadi [14] compared 17,000 matched article pairs and found that Grokipedia (v0.1) articles are longer and contain significantly fewer references per word. Their semantic similarity analysis revealed a bimodal distribution: some articles are semantically and stylistically aligned with corresponding Wikipedia articles, while others diverge sharply. Triedman and Mantzarlis [10] similarly observed that many similar Grokipedia v0.1 articles were direct reworks of high-quality Wikipedia articles, with some explicitly marked in Grokipedia as "adapted from Wikipedia" (though not all derived articles included such attribution). They further reported greater dissimilarity in subsets concerning elected officials and controversial topics, noting that Grokipedia selectively rewrites high-quality Wikipedia articles with a topical bias toward biographies, politics, society, and history. Yasseri and Mohammadi [14] identified a systematic rightward shift in the political bias of cited sources, particularly in domains related to politics, history, and religion. These studies quantify divergence at article scale using cosine similarity, BERTScore, or TF-IDF, but do not measure directional bias nor test causal drivers of divergence. We extend this literature in two ways: **(i)** we treat cities as an apolitical knowledge domain and measure directional drift as Euclidean displacement in a shared diffusion manifold in a dataset, that emhphasizes high quality data matching; **(ii)** we test temporal hypotheses (*availability & transfer bias*) with panel regressions and first-difference models.

3 Methodology

3.1 Data sampling

We compiled a sample of the 1,000 most populous cities, for each, the United States and Germany. For Germany, population figures and administrative status were obtained from the German Federal Statistical Office (Statistisches Bundesamt)[2]. For the United States, we used a publicly available compilation of the 1,000 largest incorporated places by population[3].

Each article URL was subject to a two-stage validation procedure. First, we verified that the URL resolved to an existing page. Second, disambiguation pages where filtered and resolved. All cities in the initial sample could be matched to a Wikipedia article. In contrast, Grokipedia coverage was incomplete, with missing articles particularly common among smaller German municipalities. Wikipedia's naming conventions resulted in a near-perfect matching rate (>99%). In contrast, Grokipedia exhibited frequent nonsystematic deviations in article naming. No systematic pattern explaining Grokipedia coverage gaps was evident. The validated list of URLs was then used to download article texts from both platforms. Articles from Grokipedia v0.1 and v0.2, along with their corresponding Wikipedia articles, were retrieved on October 30, 2025 (T_1) and December 8, 2025 (T_2), respectively.

Population sizes among the sampled U.S. cities ranged from 8,478,072 (New York City, NY) to 175,216 (Panama City, FL), while German cities ranged from 3,685,265 (Berlin) to 13,138 (Altötting).

In total we analysed, $N_{\text{matched}} = 1{,}387$ cities (United States: 986; Germany: 401) for which articles were available on both platforms and formed the core analytic sample. 174 articles in Grokipedia v0.1 included the notice "adapted from Wikipedia," while none in v0.2 were similarly marked. For matched cities, we additionally retrieved article edit histories. Across the Grokipedia sample, 72 cities had at least one AI-mediated edit request, while 1,306 cities had none, yielding a total of 212 requests. Edit outcomes varied, with 96 requests implemented, 77 rejected, 33 approved, and 6 remaining under review; the median request timestamp was December 1, 2025 (IQR: November 25–December 1). Grokipedia articles were substantially longer on average than their Wikipedia counterparts, containing 221 ± 100 sentences (median: 224; range: 1–1,496), compared with 140 ± 113 sentences on Wikipedia (median: 108; range: 7–804). Semantic overlap between platforms was high: the mean cosine similarity between Grokipedia and Wikipedia embeddings was 0.943 (SD = 0.052; median = 0.959). Cities in the sample varied widely in population size, spanning from small municipalities to major metropolitan areas (median population: 29,187; mean: 67,572; SD = 188,270).

3.2 Directional semantic drift

3.2.1 Semantic embedding of city articles. We computed sentence-level embeddings of all city articles using the BAAI/bge-base-en-v1.5 model [13], which generates 384-dimensional dense representations. As city articles vary substantially in length full-article embeddings collapse this structural variation, making it impossible to distinguish whether measured "drift" reflects content differences or mere length disparity. Sentence-level analysis preserves local semantic structure. Paragraph-level chunking was rejected because paragraph boundaries are inconsistent across platforms (Wikipedia uses hierarchical sections; Grokipedia uses variable paragraph breaks), introducing alignment artifacts. To retain maximal semantic information, articles were first cleaned for extraneous markup and normalized, while preserving sentence boundaries.

[1] https://github.com/nika-akin/DSD-Grokipedia
[2] Cities (all municipalities with city status) by area, population, and population density as of December 31, 2024; Statistisches Bundesamt, accessed October 30, 2025.
[3] "1000 Largest US Cities by Population" (2013 data), GitHub Gist by Miserlou; https://gist.github.com/Miserlou/11500b2345d3fe850c92, accessed October 30, 2025.

Each article was then tokenized into sentences, and we computed a length-weighted centroid embedding across all sentences, such that longer and more detailed sentences contributed proportionally more to the final vector. For each city, separate embeddings were computed for the Grokipedia and Wikipedia versions of the article at the initial time point (T_1), and in the same manner for a subset of articles from a later time point (T_2), resulting in paired representations across both platform and time. All embeddings were normalized to unit length to allow cosine similarity and diffusion-based analyses on a comparable scale.

3.2.2 Political content scoring. To quantify the political salience of city articles, we employed a zero-shot natural language classification approach using the `facebook/bart-large-mnli` model implemented in Hugging Face Transformers. This entailment-based model computes the probability that a given text expresses a predefined concept, in our case *"refugees, asylum or municipal shelter policy"*. Scores were computed independently for Grokipedia and Wikipedia articles at the initial measurement T_1 as well as T_2. The resulting political scores, representing the model's entailment probability for each article, were used as continuous predictors in both pooled and first-difference regression analyses of semantic drift.

3.2.3 Diffusion-map. Diffusion maps identify low-dimensional structure in high-dimensional data by modeling random walks across data points [2]. Unlike PCA, which finds linear axes of maximum variance, diffusion maps capture non-linear geometric structure— critical here because semantic relationships between city descriptions are unlikely to be linear [2]. We used the leading 10 eigenvectors to represent each city in a 10-dimensional latent space (`pydiffmap 0.2.0`). The symmetric Gaussian kernel bandwidth ε was set to the *median of the squared Euclidean distances* between city vectors (Scott's rule). The diffusion operator was constructed with $\alpha = 0.5$; we retained the leading $n_{\text{evec}} = 10$ eigenvectors for the final 10-D diffusion coordinates. Because diffusion coordinates are rotation-invariant, we aligned the Wikipedia and Grokipedia spaces using Procrustes analysis, which finds the optimal rotation, translation, and scaling to minimize distance between corresponding points. Pairwise Procrustes alignment (`Scipy 1.11`) registered the two 10-D spaces: first cross-platform (Grok $\mathrm{T}_1 \rightarrow$ Wikipedia T_1 = 0.53), then temporal within each platform (Grok $\mathrm{T}_1 \rightarrow \mathrm{T}_2$ = 0.99, Wiki $\mathrm{T}_1 \rightarrow \mathrm{T}_2$ = 0.99). The reported disparities (0.53 for cross-platform, 0.99 for temporal) indicate remaining misalignment after this correction—substantial between platforms, minimal over time.

The *semantic shift* of city i is the vector $(\Delta\theta_1, \Delta\theta_2)$ between its aligned Wikipedia and Grokipedia coordinates in the shared 2-D projection (diffusion components 1–2) (see Fig.1). For visual clarity arrows are coloured by the original Euclidean displacement magnitude (2–98 % range); direction therefore indicates the axis of distortion, while colour intensity encodes its strength. A 90% probability contour of the bivariate Gaussian kernel density estimate (Scott bandwidth) delineates the region of highest city density in the diffusion plane; city markers are colored by platform (Grokipedia, Wikipedia). The resulting quiver field represents city-level drift vectors, encoding both the direction and magnitude of semantic displacement between Wikipedia and Grokipedia-generated summaries.

3.2.4 Panel models. To quantify changes in semantic alignment between Grokipedia- and Wikipedia-based city representations over time, we estimated both pooled (see Eq.1) and first-difference panel regression models at the city–time level (see Eq.2). We control for the Wikipedia text length as a proxy for information volume and editorial saturation. This measure captures changes in the amount of structured knowledge available for a city. The dependent variable, *drift,* measures the cosine distance between Grokipedia and Wikipedia sentence-centroid embeddings for a given city at time T. In the pooled specification, we regress *drift* on contemporaneous *article size* (log Wikipedia text length), *Political Content* (political asylum-related content shares in Grokipedia and Wikipedia), *city population* (log), a *time* indicator, and *country* fixed effects, with standard errors clustered at the city level to account for serial correlation within cities.

$$\begin{aligned}\mathrm{Drift}_{it} = \alpha &+ \beta_1 \log(\mathrm{WikiLength}_{it}) + \beta_2 T_t \\ &+ \beta_3 \mathrm{PoliticalContent}_{it}^{\mathrm{Wiki}} + \beta_4 \mathrm{PoliticalContent}_{it}^{\mathrm{Grok}} \\ &+ \beta_5 \log(\mathrm{Population}_i) + \gamma\,\mathbf{1}_{US_i} + \varepsilon_{it}.\end{aligned} \tag{1}$$

To isolate within-city change and eliminate time-invariant city characteristics, we additionally estimate a first-differences model in which all variables are differenced between consecutive time points for each city. This specification identifies how changes in Wikipedia article size and political content on each platform are associated with changes in semantic drift, net of all fixed city attributes. All first-difference models are estimated using heteroskedasticity-robust (HC3) standard errors.

$$\begin{aligned}\Delta\mathrm{Drift}_{it} = \delta &+ \theta_1 \Delta \log(\mathrm{WikiLength}_{it}) \\ &+ \theta_2 \Delta\mathrm{PoliticalContent}_{it}^{\mathrm{Wiki}} \\ &+ \theta_3 \Delta\mathrm{PoliticalContent}_{it}^{\mathrm{Grok}} + \Delta\varepsilon_{it}.\end{aligned} \tag{2}$$

3.3 Results

Directional structure of semantic drift. For the diffusion map results, the pair (DC1, DC2) explains 33.7% of total variance and carries the largest drift signal. Figure 1 quantifies semantic distortion between AI-generated (Grokipedia) and human-curated (Wikipedia) city descriptions. After Procrustes alignment of 10-D diffusion embeddings ($\epsilon \approx 0.23$–0.32, $n_{evec} = 10$, $n = 695$ cities) arrows reveal systematic directional flows (panel 1, Fig.1). Across cities, cross-platform displacement averages 0.056 ($SD = 0.047$) diffusion units, with largest drift not in capital cities or tourist hubs. Cities in the top 5% of cross-platform drift have a mean population size of 27,182). Within-platform temporal shifts (panels 2–3) are higher and directionally consistent: Grokipedia $\mathrm{T}_1 \rightarrow \mathrm{T}_2$ mean shift 0.153 ($SD = 0.078$); Wikipedia 0.152 ($SD = 0.052$).

We quantified the directionality of semantic drift in the diffusion-map latent space by computing the mean displacement angles of cities across and within platforms. Cross-platform drift from Grok to Wikipedia exhibits a highly coherent direction, with a mean angle of -94.4° (bias-corrected accelerated 95 % CI: $[-104.5^\circ, -84.2^\circ]$), indicating a systematic shift primarily along the second diffusion coordinate. Within-platform drift over time shows distinct patterns: Grokipedia-generated summaries display a mean drift angle of -46.6° (bias-corrected accelerated 95 % CI: $[-62.1^\circ, -31.7^\circ]$),

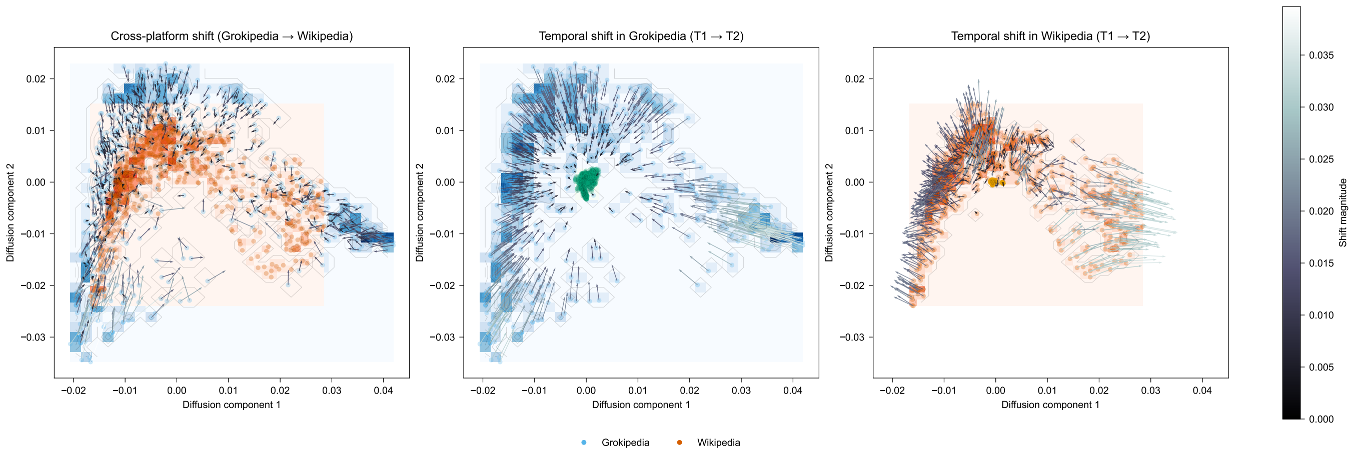

Figure 1: Semantic-drift flow fields in city representations. Each arrow denotes the displacement of a city in the diffusion-map latent space spanned by the first two diffusion coordinates (DC1–DC2). (*Left*) Cross-platform drift from Grok-generated summaries to Wikipedia articles exhibits a strong, coherent direction concentrated near -94°, indicating systematic semantic displacement across platforms. (*Center*) Within-platform drift in Grok over time follows a distinct trajectory with a mean angle of -46.6°, reflecting platform-specific semantic evolution. (*Right*) Wikipedia articles display temporal drift concentrated near 180°, consistent with gradual encyclopedic revision along an orthogonal latent dimension. Arrow orientation encodes drift direction and colour represent displacement magnitude; density contours enclose 90 % of cities per platform.

whereas Wikipedia articles exhibit a mean drift of 179.2° (bias-corrected accelerated 95 % CI: $[161.2^{\circ}, 196.6^{\circ}]$). These results demonstrate that semantic trajectories are highly structured, with cross-platform displacements following a different axis than within-platform temporal evolution. Strong angular clustering (Rayleigh $p < 10^{-19}$) confirms that city-level representations evolve along coherent latent dimensions rather than diffusing isotropically. Cities whose Grokipedia articles initially included Creative Commons (Wikipedia) licensing or attribution language but lost these references by the later snapshot (T_2) exhibited significantly greater semantic drift from Wikipedia than cities that never contained such licensing language (Two-sample Welch's t-test = 8.38, $p < 0.001$). This pattern is consistent with a transfer-bias mechanism, whereby the removal of explicit Wikipedia attribution coincides with increased divergence, indicating a new article generation instead of a Wikipedia adaptation.

Panel models. We quantify semantic divergence between Grokipedia and Wikipedia city articles using both pooled cross-sectional regressions and within-city first-difference models (Table 1). In the pooled specification, semantic drift declines sharply over time ($\beta_T = -0.184$, SE = 0.001, $p < 0.001$), indicating substantial convergence between the two platforms across measurement points. Conditional on time and city characteristics, articles from U.S. cities exhibit modestly higher drift than non-U.S. cities ($\beta = 0.022$, SE = 0.002, $p < 0.001$), consistent with stronger platform-specific framing in the U.S. context.

Political content in Grokipedia articles is negatively associated with drift in the pooled model ($\beta = -0.089$, SE = 0.012, $p < 0.001$), whereas political content in Wikipedia is not statistically distinguishable from zero. This pattern suggests that higher political salience in Grok is associated with closer semantic alignment to Wikipedia in levels, although this association may partly reflect persistent city-specific factors.

The first-difference specification, which removes all time-invariant city characteristics, reveals a complementary pattern. Increases in Wikipedia article length are strongly associated with rising semantic drift ($\theta = 0.056$, SE = 0.002, $p < 0.001$), indicating that expansions of Wikipedia content systematically shift article semantics relative to Grokipedia. Changes in political content are also predictive of divergence: a one-unit increase in Grokipedia political content is associated with a substantially larger increase in drift ($\theta = 0.203$, SE = 0.034, $p < 0.001$) than an equivalent increase in Wikipedia political content ($\theta = 0.056$, SE = 0.024, $p = 0.017$). These results indicate that temporal changes in political framing—particularly within Grokipedia—are a major driver of semantic divergence between platforms.

Taken together, the pooled and first-difference results suggest that while overall platform content converges towards Wikipedia over time, short-run shifts in article scope and political emphasis produce meaningful and asymmetric semantic divergence, with Grokipedia-specific political changes exerting the strongest effect.

4 Discussion

Our results show that semantic drift between Grokipedia and Wikipedia is structured, not stochastic noise. Two mechanisms—derived from theories of cognitive bias and domain adaptation—help explain these patterns.

Availability bias accounts for cross-sectional variation. Cities with longer Wikipedia articles exhibit greater semantic divergence from Grokipedia (not less 5.6% increase in drift per log-unit of text; Table 1). This inverts the intuition that more information

Table 1: Determinants of semantic drift between Grokipedia and Wikipedia

	(1) Pooled Levels	(2) First Differences
Political content		
Wikipedia	0.002	
	(0.011)	
Grokipedia	-0.089***	
	(0.012)	
Δ Wikipedia		0.056*
		(0.024)
Δ Grokipedia		0.203***
		(0.034)
Controls		
log(Wiki. article length)	-0.001	
	(0.002)	
Δ log(Wiki. article length)		0.056***
		(0.002)
log(Population)	-0.002	
	(0.001)	
United States (vs. GER.)	0.022***	
	(0.002)	
Time	-0.184***	
	(0.001)	
Intercept	0.273***	-0.105***
	(0.013)	(0.003)
Observations	2,774	1,782
R^2	0.812	0.402

Notes: **Column (1) reports pooled OLS estimates with city-clustered standard errors. Column (2) reports first-difference estimates based on within-city changes between the two time points. SE in parentheses. Significance levels: *** $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

improves fidelity. This association is consistent with the interpretation of **H1** that abundant documentation provides raw material for selective amplification: generative systems may emphasize, elaborate, or reframe the most detailed topics, producing directional displacement from the source. The diffusion embedding captures this as a coherent axis of variation (Fig. 1, left panel), with drift angles clustering near $-94.4°$ a signature of systematic, not random, distortion. However, we cannot rule out confounding: longer Wikipedia articles may correlate with unobserved city characteristics (economic prominence, media coverage) that independently shape Grokipedia's generation process.

Transfer bias accounts for temporal convergence. Evidence for **transfer bias (H2)** emerges from both descriptive and inferential analyses. The November 2025 platform update coincided with Grokipedia shifting toward Wikipedia by 0.11 diffusion units on average, with larger corrections for sparse-coverage cities. The first-differences regression in Table 1 confirms that this correction is largest for cities with the sparsest pre-update coverage, indicating that Grokipedia articles on cities with little available information rely heavily on Wikipedia as a main source. This heterogeneity is diagnostic: if the update improved generation uniformly, we would see parallel shifts across all cities. Instead, the pattern indicates that low-information cities were previously "anchored" to potentially low-quality sources or hallucinated content, and the new retrieval pipeline specifically remedied this deficit. Robustness checks across embedding dimensionality and kernel parameters indicate that these effects are stable to modeling choices (see Appendix Tab.2). Yet alternative explanations exist: the "new model version" may have altered generation behavior through means other than retrieval (e.g., fine-tuning, prompt changes, or filters), and our design cannot isolate the specific mechanism.

Taken together, these mechanisms interact. Grokipedia appears to interpolate between learned priors and external references: informational abundance correlates with greater drift, while platform updates correlate with convergence toward canonical sources. Whether these associations reflect deep architectural properties or specific implementation choices remains open. The diffusion-manifold framework provides a descriptive vocabulary for these dynamics—capturing static displacement structure and temporal trajectories as distinct geometric signatures. The disappearing of Wikipedia sourcing labels warrants particular caution. Articles that lost explicit Wikipedia attribution between versions became more semantically divergent, consistent with generative replacement rather than refinement. However, we cannot determine whether the model was prompted to generate new content (and this triggered both attribution removal and divergence), or whether attribution removal was a separate post-hoc policy change that preceded content regeneration.

4.0.1 Limitations. Several limitations should be acknowledged. First, our proxy for political leaning relies on asylum- and refugee-related framing. While asylum policy constitutes a salient and politically contested domain in both Germany and the United States, it represents only one dimension of municipal politics. Future work could map city articles to a broader policy taxonomy (housing, policing, climate governance) using Wikipedia's category graph or external ontologies. Second, although we analyze semantic change across platform versions, our design remains constrained by the limited availability of historical article snapshots and time-varying confounders. More generally, these limitations highlight the importance of continuous data access to emerging AI-mediated knowledge platforms. Enabling reproducible, longitudinal, and comparative research on generative knowledge systems will require infrastructures that adhere to FAIR data principles—ensuring that data are findable, accessible, interoperable, and reusable for the computational social science community [1]. Moreover, while embeddings allow us to measure cosine similarity, their validity—particularly regarding factual accuracy and neutrality—requires further investigation.

The case study introduced here opens up new research avenues how well it generalizes over domains such as contested historical events, climate policy, medical information, or electoral knowledge. As generative systems increasingly mediate access to collective memory, such tools offer a scalable, descriptive early-warning mechanism for detecting directional distortions before they become entrenched.

References

[1] Veronika Batzdorfer, Wolfgang Zenk-Möltgen, Laura Young, Alexia Katsanidou, Johannes Breuer, and Libby Bishop. 2024. Between urgency and data quality: Assessing the FAIRness of data in social science research on the COVID-19 pandemic. *Research Ethics* 20, 4 (2024), 744–763.

[2] Ronald R Coifman, Ioannis G Kevrekidis, Stéphane Lafon, Mauro Maggioni, and Boaz Nadler. 2008. Diffusion maps, reduction coordinates, and low dimensional representation of stochastic systems. *Multiscale Modeling & Simulation* 7, 2 (2008), 842–864.

[3] Selena Deckelmann. 2023. Wikipedia's value in the age of generative AI. *Wikimedia Foundation* 12 (2023).

[4] Mats Faulborn, Indira Sen, Max Pellert, Andreas Spitz, and David Garcia. 2025. Only a Little to the Left: A Theory-grounded Measure of Political Bias in Large Language Models. *arXiv preprint arXiv:2503.16148* (2025).

[5] R Alexander Knipper, Charles S Knipper, Kaiqi Zhang, Valerie Sims, Clint Bowers, and Santu Karmaker. 2025. The Bias is in the Details: An Assessment of Cognitive Bias in LLMs. *arXiv preprint arXiv:2509.22856* (2025).

[6] Zachary J McDowell. 2024. Wikipedia and AI: Access, representation, and advocacy in the age of large language models. *Convergence* 30, 2 (2024), 751–767.

[7] Aliakbar Mehdizadeh and Martin Hilbert. 2025. Epistemic Substitution: How Grokipedia's AI-Generated Encyclopedia Restructures Authority. *arXiv preprint arXiv:2512.03337* (2025).

[8] Sinno Jialin Pan and Qiang Yang. 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2009), 1345–1359.

[9] Nikhil Sharma, Q Vera Liao, and Ziang Xiao. 2024. Generative echo chamber? effect of llm-powered search systems on diverse information seeking. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–17.

[10] Harold Triedman and Alexios Mantzarlis. 2025. What did Elon change? A comprehensive analysis of Grokipedia. *arXiv preprint arXiv:2511.09685* (2025).

[11] Amos Tversky and Daniel Kahneman. 1973. Availability: A heuristic for judging frequency and probability. *Cognitive psychology* 5, 2 (1973), 207–232.

[12] Roberto Ulloa, Eve M Zucker, Daniel Bultmann, David J Simon, and Mykola Makhortykh. 2025. From prosthetic memory to prosthetic denial: Auditing whether large language models are prone to mass atrocity denialism. *AI & SOCIETY* (2025), 1–15.

[13] Shitao Xiao, Zheng Liu, Peitian Zhang, and Niklas Muennighoff. 2023. C-Pack: Packaged Resources To Advance General Chinese Embedding. arXiv:2309.07597 [cs.CL]

[14] Taha Yasseri. 2025. How Similar Are Grokipedia and Wikipedia? A Multi-Dimensional Textual and Structural Comparison. *arXiv preprint arXiv:2510.26899v2* (2025).

A Robustness check diffusion-map

Table 2: Robustness of diffusion-map embeddings

n_{evec}	ϵ	Var(DC1+DC2)	Mean angle (°)	n_{evec} used
8	scott	0.389	-112.0	8
8	0.15	0.395	-23.8	8
8	0.25	0.353	32.4	8
10	scott	0.337	-112.0	10
10	0.15	0.356	156.2	10
10	0.25	0.298	-147.6	10
12	scott	0.300	112.0	12
12	0.15	0.327	-156.2	12
12	0.25	0.260	147.6	12

We conducted a robustness check of the diffusion-map embeddings (see Tab.2) by varying the number of eigenvectors (n_{evec}) and the kernel scaling parameter (ϵ). Across parameter choices, the variance captured by the first two diffusion coordinates ranged from 0.26 to 0.40, indicating that a substantial fraction of city-level semantic variation is consistently represented. Mean drift angles varied between roughly $-156°$ and $156°$, reflecting that while the absolute direction of semantic displacement is sensitive to parameterization, the general magnitude and structure of city-level shifts remain stable.