

Online Appendix for:
The AI Democratization Paradox:
Evidence from Decentralized Knowledge Communities

Kai Zhu¹ and Dylan Walker²

¹*Bocconi University*

²*Chapman University*

Web Appendix

A Data Source

A.1 Published Translations on Wikipedia

The primary data for our analysis comprises published translations from Wikipedia’s Content Translation system. Wikimedia Foundation provides data access through Application Programming Interface (API) that lists published translations, offering comprehensive details essential for our study:

- **API Access:** The API provides lists of all published translations across languages, including details like translation IDs, translation date, source and target titles, source

and target languages, URLs of the articles, and translation statistics indicate the extent of human versus machine translation.

- **Parallel Text Dump:** The tool also allows access to parallel text data, encompassing the source content and the final user-translated content.

This data is provided under an open license and represents an ever-improving resource for cross-linguistic research and machine translation development. For more information, refer to the MediaWiki page: [Content Translation - Published Translations](#)

A.2 Wikipedia Knowledge Gap Index

The Knowledge Gap Index developed by the Wikimedia Foundation is instrumental in quantifying knowledge disparities across various dimensions on Wikipedia. The dataset is structured to provide a clear mapping of content gaps across languages

- **Content Mapping:** The Knowledge Gap Index employs a content mapping system, categorizing Wikipedia articles into specific domains and themes. This system allows us to assess disparities in content availability and representation across diverse topics and regions.
- **Gender Gap Measurement:** the Knowledge Gap Index addresses the Gender Gap. This measure assesses the balance in representation of different genders across Wikipedia articles. It aids in quantifying and understanding the skewness in gender portrayal,
- **Geographic Gap Measurement:** Another useful aspect of the Knowledge Gap Index is its focus on the Geographic Gap. This measurement evaluates the extent of content coverage pertaining to different geographic regions.
- **Open Access and Continuous Updates:** The Knowledge Gap Index is publicly accessible and is regularly updated, reflecting the dynamic nature of content creation on Wikipedia.

The Knowledge Gap Index offers a detailed perspective on the existing content distribution and the potential areas where machine translation can have a significant impact. For further details about the dataset and its applications, please refer to the Wikimedia Meta-Wiki page on the [Knowledge Gap Index](#) and its [Geographic Gap Measurement](#).

B Methodology

B.1 Classifying Topics in Wikipedia Articles

Our approach to classifying article topics across different languages in Wikipedia builds on the language-agnostic methodology developed by [Johnson et al. \(2021\)](#). This method involves several steps:

1. **Topic Taxonomy:** Utilizing a taxonomy of 64 topics derived from WikiProject tags, encompassing a broad range of subjects from science and technology to arts and culture.
2. **Link-Based Representation:** Articles are represented by Wikidata items of contained links, facilitating a language-independent content representation.
3. **Mapping Articles to Topics:** Machine learning algorithms map articles to the pre-defined topics based on link similarities.
4. **Ensuring Cross-Linguistic Consistency:** The focus on link-based features rather than textual content assures consistent classification across languages.
5. **Validation:** Classification accuracy is validated and refined through comparison with a manually classified article sample.

This methodology enables a comprehensive and consistent analysis of article topics across multiple language editions of Wikipedia, overcoming the limitations of language-specific models and ensuring uniform topic classification.

B.2 Extracting Attributes from Wikidata

We utilized Wikidata to extract demographic and geographical data for our study. We used the Wikidata Query Service, which allows for sophisticated and precise queries, to retrieve the required information efficiently and accurately

- **Biographical Articles:** Information such as gender, sexual orientation, and ethnicity of individuals was systematically extracted from their respective Wikidata entries. This process involved querying Wikidata's structured data to obtain these specific attributes for each biographical article.
- **Geographical Articles:** For articles pertaining to geographical locations, we extracted location data, including coordinates and administrative hierarchies. This was achieved by querying the corresponding Wikidata items linked to each geographical article.

This methodology ensured a standardized and accurate extraction of demographic and geographical information, useful for our analysis.

C Robustness for Aggregate Impact of Google Translate Deployment

C.1 Primary Design: Alternative DID Estimator

C.1.1 More Information about Matrix Completion Method Estimation

The matrix completion method provides additional validation of our main findings on Google Translate’s impact. As shown in [Table A.1](#), The matrix completion method approach estimates an Average Treatment Effect on the Treated (ATT) of 71.2 additional monthly translations ($p < 0.01$), with a 95% confidence interval of [38.4, 104.1]. This result aligns with our primary difference-in-differences estimates and the two-way fixed effects model, providing robust evidence that Google Translate integration substantially increased translation volume across Wikipedia languages.

Table A.1: Matrix Completion Method: Estimated Impact of Google Translate Integration on Monthly Translation Volume

ATT	S.E.	CI.lower	CI.upper	P.value
71.2	16.8	38.4	104.1	0.0

Note: This table presents the Average Treatment Effect on the Treated (ATT) estimated using the matrix completion method. The ATT of 71.2 indicates the average increase in monthly translations attributable to Google Translate integration. S.E. represents standard error, CI shows 95% confidence interval bounds, and P.value indicates statistical significance.

C.1.2 Two-way Fixed Effect

This model estimates the Average Treatment Effect (ATE) of Google Translate’s integration on the volume of translated articles. By incorporating both language and time fixed effects, this approach controls for time-invariant language characteristics and common temporal shocks across languages. The statistically significant treated coefficient of 61.4 ($p < 0.01$)

in [Table A.2](#) reinforces the substantial impact of Google Translate on content translation activities in Wikipedia, while accounting for potential confounding factors through the two-way fixed effect structure.

Table A.2: Two-way Fixed Effects Analysis of Google Translate Integration Impact

Dependent Variable: Model:	Number of Translated Articles (1)
Google Translate Integration	61.4*** (14.9)
<i>Fixed-effects</i>	
Language	Yes
Year-Month	Yes
<i>Fit statistics</i>	
Observations	6,336
R ²	0.77

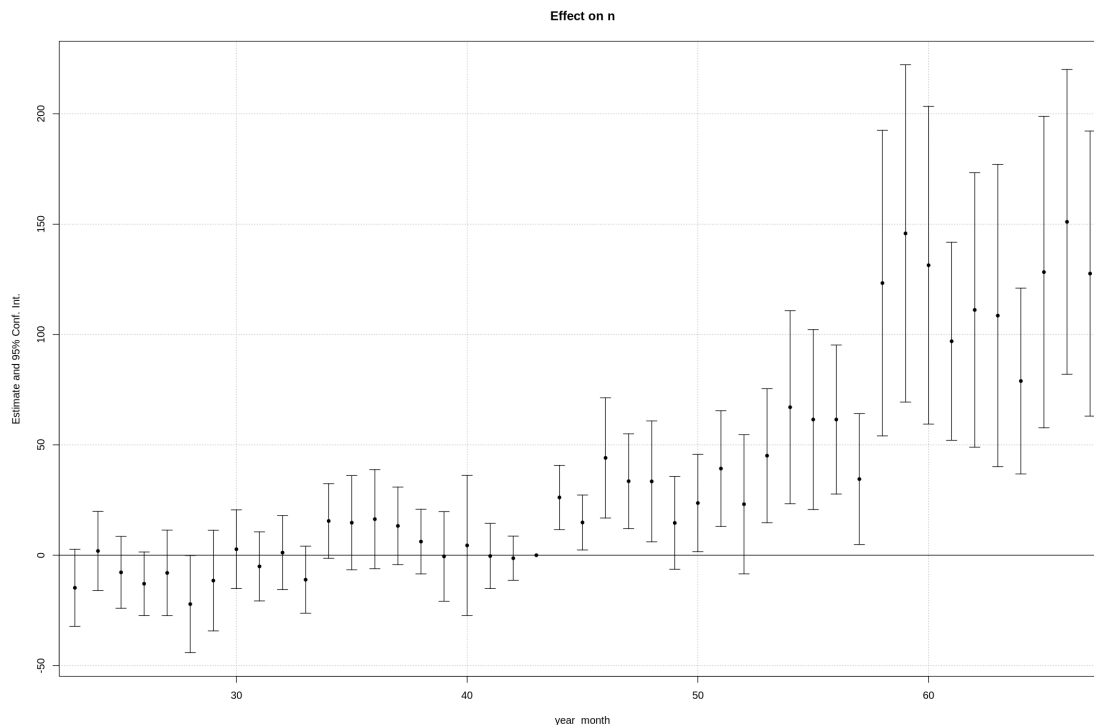
Clustered (Language) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table presents results from a two-way fixed effects model examining the impact of Google Translate integration on monthly translation volume. The coefficient of 61.4 represents the average increase in monthly translations after integration, controlling for both language-specific and temporal effects. Standard errors are clustered at the language level.

C.1.3 Sun and Abraham Weighted Interactions

The Sun and Abraham Weighted Interactions method addresses potential biases in staggered difference-in-differences designs by properly accounting for treatment effect heterogeneity. This approach computes cohort-specific treatment effects and combines them using weights that reflect the relative size of each cohort. The method highlights the temporal dynamics of the treatment effect, indicating a sustained positive impact of Google Translate over time on the translation volume. By explicitly modeling treatment effect heterogeneity across adoption cohorts, this approach provides more reliable estimates of the dynamic treatment effects.

Figure A.1: Dynamic Treatment Effects of Google Translate Integration Using Sun and Abraham Method



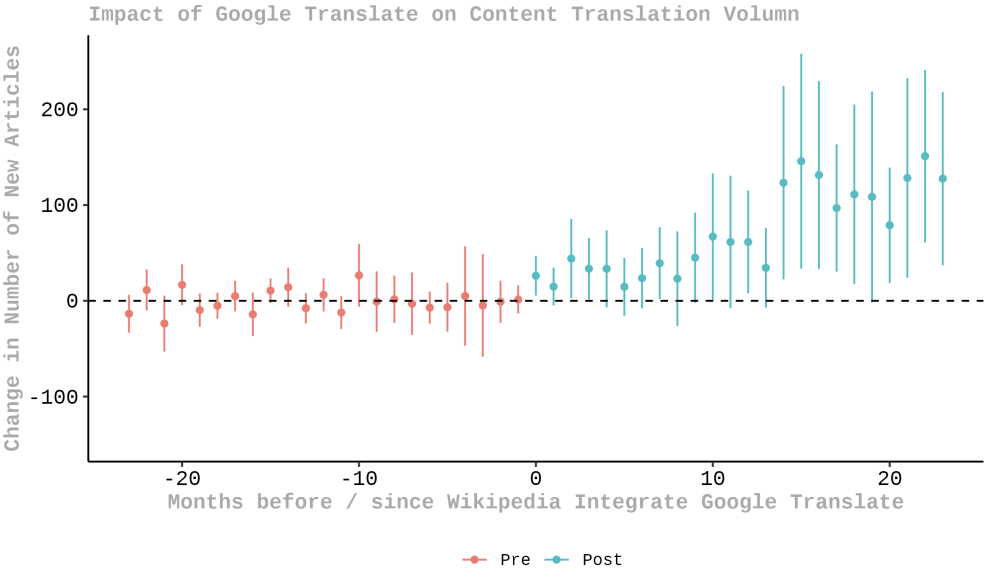
Note: This figure displays the dynamic treatment effects estimated using the Sun and Abraham weighted interactions method. The vertical axis shows the estimated effect on monthly translation volume, while the horizontal axis represents time relative to Google Translate integration. The solid line indicates point estimates, and the shaded area represents 95% confidence intervals. The results demonstrate persistent positive effects of integration across different time periods, accounting for heterogeneous treatment effects across adoption cohorts.

C.1.4 Callaway and Sant’Anna Estimators

We further assess the robustness of our findings using the Callaway and Sant’Anna Estimators, which are specifically designed for settings with staggered treatment adoption and heterogeneous treatment effects. This methodology offers several advantages: it allows for treatment effect heterogeneity both across groups and over time, remains valid under parallel trends assumptions, and is robust to negative weights that can arise in two-way fixed effects models. The estimators compute group-time average treatment effects and aggregate them into overall effects using carefully chosen weights. The Callaway and Sant’Anna Estimators

confirm the robustness of our main findings, showing consistent positive effects of Google Translate’s integration on translation volume across various Wikipedia language editions.

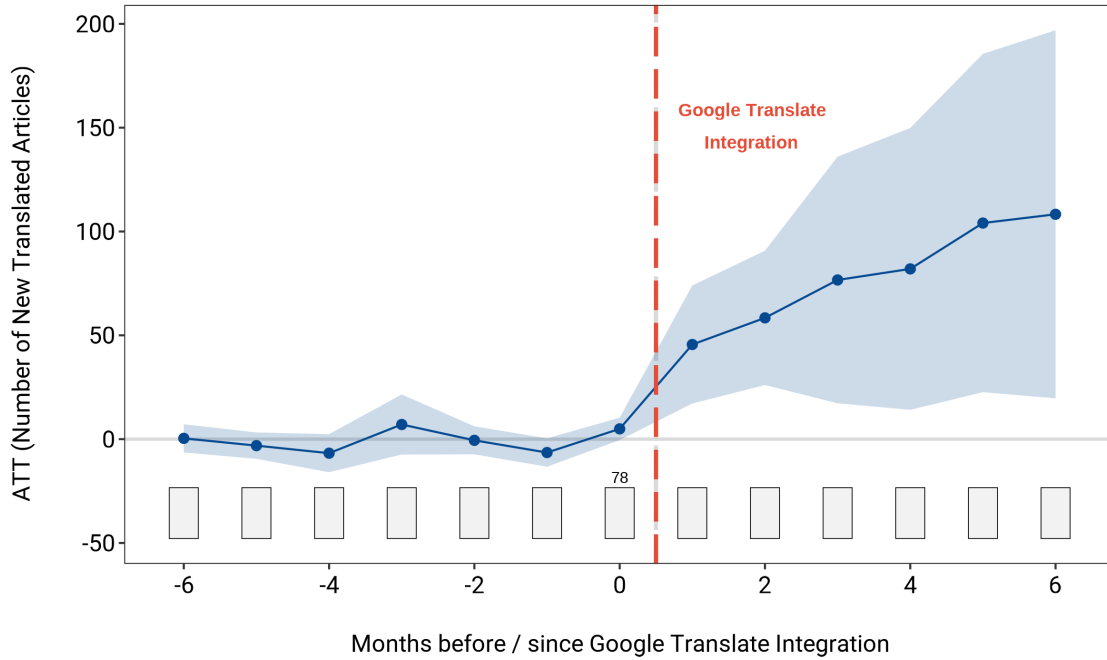
Figure A.2: Group-Time Average Treatment Effects Using Callaway and Sant’Anna Estimators



Note: This figure presents the estimated group-time average treatment effects using the Callaway and Sant’Anna method. The vertical axis shows the estimated effect on monthly translation volume, while the horizontal axis represents time relative to Google Translate integration. The solid line shows point estimates, and the shaded area represents 95% confidence intervals. The results demonstrate robust positive effects across different adoption groups and time periods, avoiding potential biases from negative weighting in traditional two-way fixed effects models.

C.2 Alternative Identification: Time-Shift DID Design

Figure A.3: Effect of Google Translate Integration on Monthly Translation Volume



Note: Event study plot showing changes in translation volume relative to Google Translate integration ($t=0$). The black line represents point estimates with 95% confidence intervals (shaded area).

D Relative Effects Analysis: Additional Tests of the Matthew Effect

Our main analyses document substantial heterogeneity in absolute treatment effects across language communities, with well-resourced editions capturing disproportionately larger gains. A natural concern is whether these absolute differences could be partially mechanical—larger communities might naturally produce more translations simply due to their greater scale. In this appendix, we present a series of complementary analyses examining *relative* effects, which provide a more stringent test of the Matthew effect. If well-resourced communities gain more even after adjusting for their larger base, this strengthens the case that structural advantages—not mere scale—drive the concentration of benefits.

D.1 Fixing the Source Language: English-Only Translations

One potential confound in our heterogeneity analysis is that variation in treatment effects could reflect differences in *source* languages rather than *target* community resources. If well-resourced target languages happen to draw from higher-quality or more extensive source pools, this could mechanically produce larger gains without implying any Matthew effect on the target side.

As documented in Section 7.1, English Wikipedia dominates as a translation source, with its share increasing from 68% to 81% post-integration. Meanwhile, virtually no content flows *into* English Wikipedia, consistent with its editorial policies restricting translated content. This asymmetry provides an opportunity for a cleaner test: by restricting our sample to translations originating from English only, we hold constant the source language and isolate variation in how different target-language communities absorb content from the *same* origin.

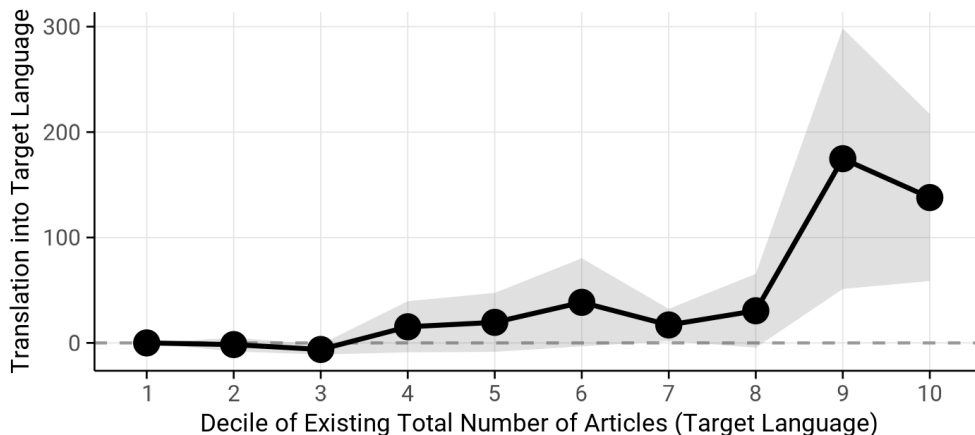
Figure A.4 presents the results. The gradient pattern persists: among translations from English, well-resourced target languages still capture disproportionately larger gains. This confirms that the Matthew effect reflects genuine differences in target-community resources—editorial capacity, absorptive infrastructure, reader networks—rather than an artifact of source-language variation.

D.2 Scaling by Target Language Resource Base

Our absolute-effects findings document that well-resourced communities gain more in raw volume, which is itself informative about the “rich get richer” phenomenon. However, a more stringent test asks whether well-resourced communities also gain more *relative to their size*. After all, larger communities have more editors, more existing content to build upon, and more organizational infrastructure, so larger absolute gains could be partially mechanical. If the Matthew effect is genuine, we should observe disproportionate gains even after adjusting for baseline resource differences.

To test this, we re-estimate our difference-in-differences models with the outcome variable

Figure A.4: Heterogeneous Treatment Effects: English-Source Translations Only



Notes: Treatment effects estimated on the subset of translations where English Wikipedia is the source language. Restricting to a common source controls for variation in source-language characteristics (content availability, quality, breadth), isolating differences in target-language capacity to absorb and integrate translated content. The persistent gradient confirms that the Matthew effect reflects target-community resources, not source-language composition. Error bars represent 95% confidence intervals.

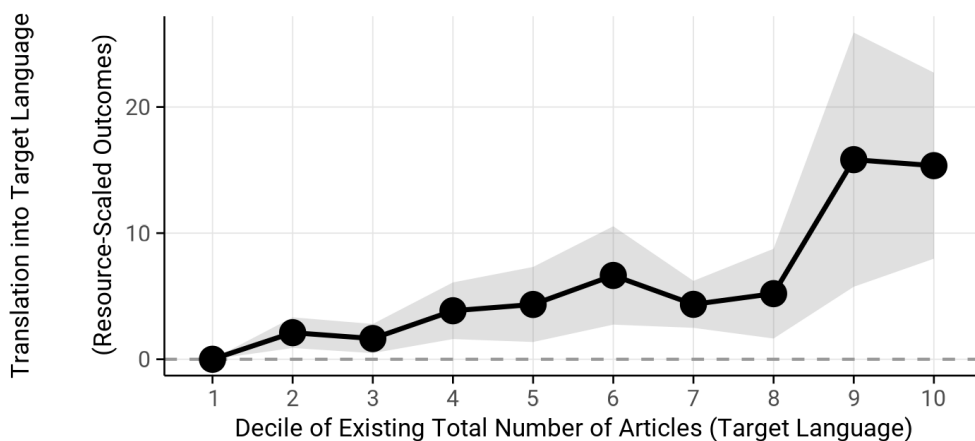
scaled by each target language’s resource base. Specifically, we divide monthly translated articles by the natural logarithm of total existing articles in each language edition. This scaling transforms our outcome from “how many articles were translated” to “how many articles were translated relative to what the community already had”—a direct test of whether proportional gains also favor well-resourced communities.

We use the natural logarithm to account for the extremely skewed distribution of resources across Wikipedia language editions. English Wikipedia has over 6 million articles, while smaller editions have fewer than 100,000—a difference of nearly two orders of magnitude. The log transformation compresses this variation and yields more interpretable comparisons across the full range of community sizes.

Figure A.5 presents the results. The gradient pattern not only persists but remains striking: top-tier communities see gains more than 10× larger than bottom-tier communities, even after adjusting for their larger base. This confirms that the Matthew effect operates

on relative shares, not just absolute volumes. Well-resourced communities capture disproportionate benefits whether we measure in raw articles or in proportion to their existing content.

Figure A.5: Heterogeneous Treatment Effects with Resource-Scaled Outcome



Notes: Treatment effects from difference-in-differences estimation where the dependent variable is monthly translated articles scaled by the log of total existing articles in each language edition. The gradient pattern observed in absolute terms persists: well-resourced communities capture disproportionately larger gains even after accounting for their larger resource base. Error bars represent 95% confidence intervals.

D.3 Heterogeneity by Pre-Existing Translation Dependency

The previous analyses scaled by resource base and fixed the source language. We now examine heterogeneity from a different angle: how do effects vary by communities’ *baseline reliance* on translation? This approach leverages the pre-existing structural differences documented in Section 4.2, where we showed that smaller, resource-constrained editions depend far more heavily on translation (15–20% of total content) than larger editions (1–5%).

This analysis provides a theoretically informative test. If AI deployment narrows gaps, we would expect high-dependency communities—those most reliant on translation and with the greatest apparent need—to see the largest gains. The technology would help them “catch up” to better-resourced editions. Conversely, if the Matthew effect holds, we would expect

the opposite: low-dependency communities with greater absorptive capacity would leverage the technology more effectively, widening rather than narrowing existing disparities.

We categorize languages into terciles based on their pre-intervention translation dependency (translated articles as a share of total content): high dependency (15–20%), medium dependency (5–15%), and low dependency (1–5%). We then estimate a heterogeneous difference-in-differences model interacting treatment status with dependency tercile indicators.

Table A.3 reports the results. Large editions with *low* baseline dependency—those least reliant on translation before the intervention—saw the largest gains (109.7 additional articles per month, $p < 0.01$). Medium-dependency communities saw moderate gains (28.5 articles), while high-dependency communities—small editions already heavily reliant on translation—saw only modest, statistically insignificant increases (3.3 articles).

Table A.3: Heterogeneous Treatment Effects by Pre-Intervention Translation Dependency

	Translated Articles
Treated × High Dependency	3.3 (6.4)
Treated × Medium Dependency	28.5 (23.7)
Treated × Low Dependency	109.7*** (41.0)
Language Fixed Effects	Yes
Year-Month Fixed Effects	Yes
Observations	6,048
R ²	0.773
Within R ²	0.043

Notes: Clustered standard errors (by language) in parentheses. Dependency terciles based on pre-intervention (2017–2018) ratio of translated articles to total articles. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

This pattern provides strong support for the Matthew effect. Communities with greater absorptive capacity—more editors to review content, stronger organizational infrastructure, larger reader bases to justify the effort—leveraged AI to substantially expand their translation activity. Meanwhile, resource-constrained communities that already depended heavily

on translation saw limited additional benefit from the technological improvement. Rather than narrowing the gap in translation intensity, AI deployment widened it.

D.4 Summary

The analyses in this appendix complement our main findings by examining relative effects from multiple angles. Three key conclusions emerge:

1. **The Matthew effect is not a source-language artifact.** When we restrict to English-source translations, holding constant the origin of content, the gradient pattern persists. Target-community resources, not source-language composition, drive the heterogeneity.
2. **The Matthew effect operates on relative shares, not just absolute volumes.** After scaling by resource base, well-resourced communities still capture gains more than $10\times$ larger than resource-constrained communities. The concentration of benefits is not merely mechanical.
3. **AI widened rather than narrowed translation dependency gaps.** Communities least reliant on translation before the intervention benefited most, while those most dependent saw minimal gains. The technology amplified rather than alleviated pre-existing structural disparities.

Together with the content-category benchmarking analysis in Section 7.2—which shows that relative patterns in *what* gets translated (e.g., female biographies at $2\times$ the expected rate) reveal meaningful variation obscured by absolute volumes—these findings provide robust, multi-faceted evidence for the AI Democratization Paradox.

E Additional Supporting Results

The results presented in this section provide detailed model estimates and statistical evidence supporting the key findings discussed in the main paper. While the main text focuses on the high-level implications and interpretation of our analyses, here we present tables showing the full model specifications, robustness checks, and additional statistical tests.

E.1 Article Quality Analysis

We analyze the evolution of article quality between 2018 and 2019. The results show a modest but statistically significant improvement in article quality scores over this period, with the mean quality score increasing from 0.444 in 2018 to 0.466 in 2019. The distribution of scores, as visualized through both box plots and density plots, reveals that while there is considerable variation in article quality within each year, there is a consistent upward shift in the overall quality distribution from 2018 to 2019.

Figure A.6: Distribution of Article Quality Scores by Year (2018-2019)



Note: This figure shows the distribution of article quality scores comparing 2018 and 2019. The box plots display the median, quartiles, and range of quality scores, with the underlying data points shown as a density plot in gray. The red diamond indicates the mean quality score for each year (0.444 in 2018 and 0.466 in 2019). The extremely low p-value ($<2e-16$) suggests a statistically significant difference between the two years' distributions. The plot demonstrates a slight improvement in average article quality from 2018 to 2019.

E.2 Language Community Characteristics

Using the difference-in-differences estimator described in the Data and Methods section, we examine how the impact of Google Translate varies across different dimensions of language community characteristics. The results in [Table A.4](#) show heterogeneous effects across language tiers based on article count, editor count, view count, and Google Translate quality. When languages are categorized by article count, medium-tier languages saw an increase of 78.3 articles per month and high-tier languages experienced an increase of 181.3 articles compared to the baseline, both statistically significant. Similar patterns emerge when languages are categorized by editor and view counts, with high-tier languages showing the largest gains (256.6 articles for both metrics). Interestingly, when considering Google Translate quality, only medium-tier languages show a significant positive effect of 49.4 articles, while high-tier languages show no significant increase. These findings suggest that community size and engagement are crucial factors in leveraging machine translation technology, potentially more so than the underlying quality of the translation system.

Table A.4: Heterogeneous Effects of Google Translate Integration Across Language Community Characteristics

Dependent Variable:	Number of New Translations			
Language Tier Based On:	Article Count	Editor Count	View Count	GT Quality
Model:	(1)	(2)	(3)	(4)
<i>Variables</i>				
GT Integration \times Low-tier	13.3 (9.5)	0.6 (5.8)	4.3 (6.9)	-10.2 (20.5)
GT Integration \times Medium-tier	78.3** (30.2)	56.6*** (19.9)	52.9*** (19.8)	49.4** (23.8)
GT Integration \times High-tier	181.3*** (59.7)	256.6*** (73.0)	256.5*** (73.0)	-19.9 (29.0)
<i>Fixed-effects</i>				
Language	Yes	Yes	Yes	Yes
Year-Month	Yes	Yes	Yes	Yes
<i>Fit statistics</i>				
Observations	6,048	6,048	6,048	3,744
R ²	0.777	0.790	0.789	0.722
Within R ²	0.059	0.112	0.110	0.015

Clustered (Language) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table presents difference-in-differences estimates examining how Google Translate’s impact varies across language community characteristics. Each column represents a different categorization of languages: by article count, editor count, view count, and translation quality. The coefficients show the differential impact across low, medium, and high-tier languages, with standard errors clustered at the language level. The results demonstrate that community size and engagement significantly influence the effectiveness of machine translation integration.

E.3 Ruling Out Google Translate Quality as an Alternative Explanation

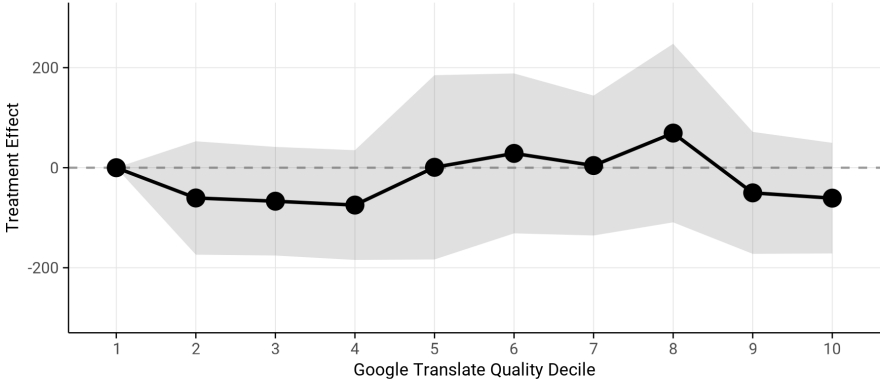
One potential concern is that Google Translate quality varies across language pairs. If well-resourced communities happen to be served by higher-quality translation, the observed gradient could reflect translation quality rather than complementary resources.

To address this, we examine whether treatment effects vary systematically with transla-

tion quality. We measure quality using BLEU scores from established machine translation benchmarks and partition languages into deciles accordingly. Figure A.7 presents the results. In stark contrast to the pronounced gradient observed for community resources (??), treatment effects do not vary systematically with translation quality. The pattern is essentially flat, with point estimates fluctuating around zero and no discernible trend.

This null result indicates that the Matthew effect is mainly driven by differences in complementary resources—editorial capacity, existing knowledge bases, and reader networks—rather than by variation in translation quality. Communities benefit differentially not because they receive better translations, but because they possess greater capacity to review, refine, and integrate translated content at scale.

Figure A.7: Heterogeneous Treatment Effects by Google Translate Quality



Notes: Heterogeneous difference-in-differences estimates by deciles of Google Translate quality (BLEU scores). Decile 1 = lowest quality; Decile 10 = highest. Unlike the gradient for community resources, treatment effects do not vary with translation quality, ruling out differential quality as a driver of the Matthew effect. Shaded bands denote 95% confidence intervals.

Table A.5: Differential Impact of Google Translate Integration Across Content Domains

Dependent Variables: Model:	Culture (1)	Geography (2)	STEM (3)	History and Society (4)
GT Integration	25.2*** (5.9)	21.4*** (5.4)	7.9*** (2.1)	5.6** (2.3)
<i>Fixed-effects</i>				
Language	Yes	Yes	Yes	Yes
Year-Month	Yes	Yes	Yes	Yes
<i>Fit statistics</i>				
Observations	6,336	6,336	6,336	6,336
R ²	0.76	0.64	0.62	0.58
Pre-integration Mean	20.8	12.5	7.2	4.3

Clustered (Language) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table presents difference-in-differences estimates showing how Google Translate’s impact varies across different content domains. The coefficients represent the average monthly increase in translations for each content type after integration. Standard errors are clustered at the language level, and pre-integration means are provided for context. The results reveal stronger effects for cultural and geographical content compared to technical and historical topics.

E.4 Content Domain Analysis

Our difference-in-differences analysis reveals varying impacts across content domains, as shown in Table A.5. Culture and Geography articles saw the largest increases (25.2 and 21.4 articles per month respectively), while STEM and History/Society topics experienced more modest growth (7.9 and 5.6 articles). This pattern suggests that Google Translate may be more effective at handling cultural and geographical content compared to technical or historical material.

E.5 Gender Analysis

The difference-in-differences estimates in [Table A.6](#) reveal persistent gender disparities in biographical translations after Google Translate integration. Male biographies increased by 7.2 articles monthly compared to 3.8 for female biographies, while non-binary biographies saw minimal growth (0.06). This suggests that machine translation may reinforce existing gender biases in content creation.

Table A.6: Differential Impact of Google Translate Integration on Biographical Content by Gender

Dependent Variables: Model:	Male Biography (1)	Female Biography (2)	Non-binary Biography (3)
GT Integration	7.2*** (2.5)	3.8*** (1.3)	0.06*** (0.02)
<i>Fixed-effects</i>			
Language	Yes	Yes	Yes
Year-Month	Yes	Yes	Yes
<i>Fit statistics</i>			
Observations	6,288	6,288	6,288
R ²	0.72317	0.54092	0.26982
Pre-integration Mean	12.5	7.8	0.06

Clustered (Language) standard-errors in parentheses

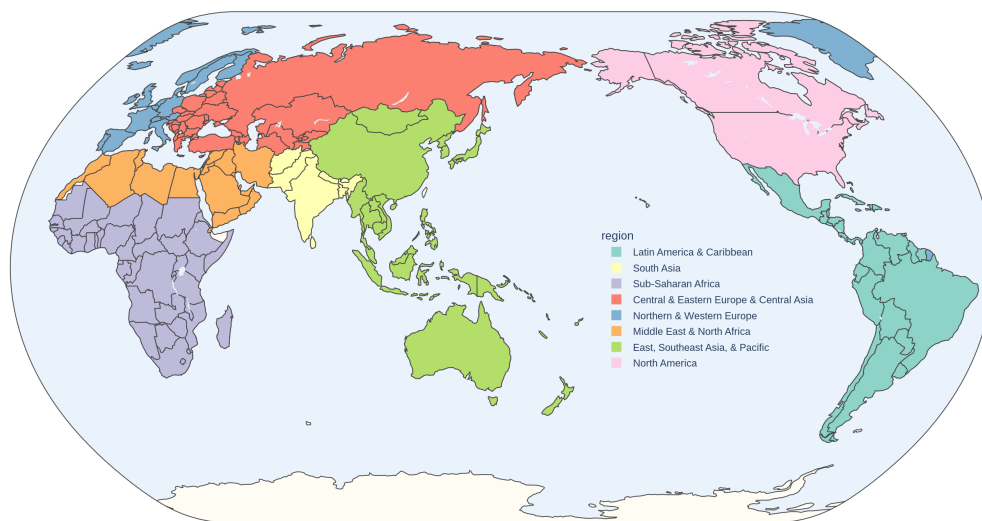
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table presents difference-in-differences estimates showing how Google Translate integration affects the translation of biographical articles across gender categories. The coefficients represent the average monthly increase in translations for each gender category after integration. Standard errors are clustered at the language level, and pre-integration means are provided for context. The results reveal a persistent gender gap, with male biographies receiving substantially more translations than female or non-binary biographies.

E.6 Geographic Analysis

Figure A.8 illustrates the geographical framework we use to analyze regional patterns in Wikipedia translations. By adopting the Wikimedia Foundation’s official regional classification system, which divides the world into eight distinct regions, we can systematically examine how Google Translate’s impact varies across different geographical and cultural contexts. This standardized regional breakdown enables us to identify potential disparities in translation activity and technological adoption across diverse global regions.

Figure A.8: Wikimedia Foundation’s Regional Classification System for Global Content Analysis



Note: This world map illustrates the eight regional classifications used by the Wikimedia Foundation for content analysis: Latin America & Caribbean (LAC), South Asia (SA), Sub-Saharan Africa (SSA), Central & Eastern Europe & Central Asia (CEECA), Northern & Western Europe (NWE), Middle East & North Africa (MENA), East/Southeast Asia & Pacific (ESEAP), and North America (NA). These regions form the basis for analyzing geographical patterns in Wikipedia translation activity and content distribution.

Using our difference-in-differences framework, we examine geographical coverage across eight major world regions defined by the Wikimedia Foundation (shown in [Figure A.8](#)). The results in the following table show that Central & Eastern Europe & Central Asia (CEECA) and East/Southeast Asia & Pacific (ESEAP) regions saw the largest increases in geographical article translations (8.6 and 7.3 articles monthly), while regions like Middle East & North Africa (MENA) and Sub-Saharan Africa (SSA) experienced minimal growth (0.3 and 0.4 articles). This regional disparity suggests that machine translation may be reinforcing existing geographical biases in content coverage.

Table A.7: Effects of Google Translate Integration on Monthly Geography Translations by Regions

DVs:	ESEAP	SA	MENA	SSA	NA	SCA	CEECA	NWE
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Variables</i>								
GT Integration	7.3*** (2.4)	5.3*** (1.8)	0.3** (0.1)	0.4** (0.2)	1.7*** (0.5)	1.3** (0.5)	8.6*** (2.6)	5.4*** (1.4)
<i>Fixed-effects</i>								
Language	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
<i>Fit statistics</i>								
Observations	6,336	6,336	6,336	6,336	6,336	6,336	6,336	6,336
R ²	0.49	0.44	0.54	0.26	0.47	0.24	0.55	0.60
Pre-integration Mean	4.7	2.8	0.2	0.2	1.0	0.6	5.0	3.4

Clustered (Language) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table presents difference-in-differences estimates showing the impact of Google Translate integration on monthly article translations across eight Wikimedia regions. ESEAP: East/Southeast Asia & Pacific; SA: South Asia; MENA: Middle East & North Africa; SSA: Sub-Saharan Africa; NA: North America; SCA: Latin America & Caribbean; CEECA: Central & Eastern Europe & Central Asia; NWE: Northern & Western Europe. The coefficients represent the average monthly increase in translations for each region after integration. Standard errors are clustered at the language level, and pre-integration means are provided for context.

E.7 Geographical Wikipedia Articles by Continents

The [Table A.8](#) presents the difference-in-differences analysis results, capturing the impact of Google Translate’s roll-out on the volume of articles across different continents. The results show that regions like Asia and Europe experienced a significant increase in article volume, with coefficients of 9.0 and 7.8, respectively, suggesting a more pronounced adoption of machine translation in these areas. America and Africa also saw increases, albeit at a lower magnitude, while Oceania demonstrated a modest but significant rise. These findings reflect the variable impact of machine translation across diverse geographical contexts.

Table A.8: Impact of Google Translate Roll-out on Article Volume by Geography Regions

Dependent Variables: Model:	America (1)	Europe (2)	Asia (3)	Oceania (4)	Africa (5)
Google Translate Integration	3.0*** (1.0)	7.8*** (2.0)	9.0*** (3.3)	0.5*** (0.2)	1.5** (0.6)
<i>Fixed-effects</i>					
Language	Yes	Yes	Yes	Yes	Yes
Year-Month	Yes	Yes	Yes	Yes	Yes
<i>Fit statistics</i>					
Observations	6,336	6,336	6,336	6,336	6,336
R ²	0.42505	0.58168	0.42917	0.29379	0.45671
Within R ²	0.00453	0.00887	0.00744	0.00226	0.00466

Clustered (Language) standard-errors in parentheses

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Note: This table presents difference-in-differences estimates showing the impact of Google Translate integration on monthly article translations across five continental regions. The coefficients represent the average monthly increase in translations for each continent after integration. Standard errors are clustered at the language level. The results reveal substantial variation in the effect across continents, with Asia and Europe showing the largest increases in translation volume.

F Analysis of Potential Displacement Effects

This section examines potential displacement effects of Google Translate integration on non-translated content creation and consumption. We analyze whether increased machine-translated content affects the production and readership of native content through resource reallocation or shifts in reader attention.

Data Collection We constructed a comprehensive panel dataset tracking monthly article creation across both translated and non-translated content in treated and control Wikipedia language editions from 2017 to 2020. This dataset captures content creation through all methods across more than 100 Wikipedia language communities, extending beyond articles created through the Content Translation system. Additionally, we collected monthly aggregated pageview statistics sourced from billions of records across Wikipedia language editions to analyze content consumption patterns.

Content Creation Analysis Using this rich dataset, we employed two model specifications examining:

1. The relationship between translated and non-translated content using fixed-effects regression controlling for language and time fixed effects
2. The impact of Google Translate integration on non-translated article creation using difference-in-differences, where the treatment effect is measured as the interaction between treatment group and post-integration period

For each specification, we analyzed both standardized article counts and log-transformed article counts to ensure robustness of our findings.

Table A.9: Analysis of Translated vs. Non-Translated Article Creation

Variables	Dependent Variable: Non-Translated Article Creation			
	Fixed Effects Model		Difference-in-Differences	
	Standardized	Log	Standardized	Log
	(1)	(2)	(3)	(4)
Number of Translated Articles	-0.05 (0.04)	0.04 (0.03)		
Google Translate Integration			-0.18 (0.18)	-0.30 (0.18)
Language Fixed Effects	Yes	Yes	Yes	Yes
Year-Month Fixed Effects	Yes	Yes	Yes	Yes
Observations	5,856	5,856	5,856	5,856
R ²	0.20	0.84	0.20	0.85

Notes: Table reports regression results examining whether increased translated content displaces non-translated article creation. Standardized variables have mean zero and standard deviation one. Standard errors (in parentheses) are clustered at the language level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Content Consumption Analysis To examine potential displacement effects in article consumption, we analyzed the relationship between translated and non-translated article pageviews using similar methodological approaches:

1. A fixed-effects regression examining the direct relationship between translated and non-translated article pageviews
2. A difference-in-differences specification analyzing how Google Translate integration affects non-translated article pageviews

The analysis incorporated both standardized and log-transformed pageview counts to ensure robust findings across different specifications.

Table A.10: Analysis of Translated vs. Non-Translated Article Consumption

Variables	Dependent Variable: Non-Translated Article Pageviews			
	Fixed Effects Model		Difference-in-Differences	
	Standardized	Log	Standardized	Log
	(1)	(2)	(3)	(4)
Number of Translated Pageviews	0.12*** (0.03)	0.18*** (0.05)		
Google Translate Integration			0.02 (0.012)	0.03 (0.04)
Language Fixed Effects	Yes	Yes	Yes	Yes
Year-Month Fixed Effects	Yes	Yes	Yes	Yes
Observations	5,856	5,856	5,856	5,856
R ²	0.99	0.99	0.99	0.99

Notes: Table reports regression results examining whether increased translated content affects pageviews of non-translated articles. Standardized variables have mean zero and standard deviation one. Standard errors (in parentheses) are clustered at the language level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Our analysis reveals several findings. For content creation, we find no significant negative effects across all specifications, indicating that increased translated content does not displace native content production. For content consumption, we observe a positive and significant relationship between translated and non-translated article pageviews in the fixed effects model (columns 1-2), suggesting complementarity rather than substitution. The difference-in-differences analysis (columns 3-4) shows no significant negative effect of Google Translate integration on non-translated article pageviews. When examining combined normalized pageviews of source and target articles, we observed an increase from a baseline of 1.0 to approximately 1.5, indicating that translations attract new readers while maintaining

the readership of original content.

These findings demonstrate that the integration of Google Translate expands Wikipedia's content ecosystem rather than creating displacement effects. Machine translation appears to complement rather than substitute for native content creation and consumption, contributing to the overall growth of Wikipedia's knowledge base across languages.

G Increased Participation vs. Editor productivity

We examine individual editor contributions following the integration of Google Translate and examine the dual forces of participation and productivity among Wikipedia editors. This analysis aims to dissect the increase in content production and discern whether it stems from a growth in the editor base or improved editor productivity.

For this investigation, we collect editor IDs for each article created using the Content Translation system in our sample, enabling us to pinpoint the respective editors behind translated articles¹. We used this information to create a language-month panel and calculate the number of new editors who used translation to create an article for the first time, the number of unique active editors contributing at least one article in a specific language and month, and the productivity of each editor, measured as the total articles created in the given month. For simplicity and robustness, we present results from analysis focused on language editions that integrated Google Translate in January 2019 and utilize two-way fixed effect models to analyze the editor panels in [Table A.11](#).

As detailed in [Table A.11](#), our analysis reveal a 49% growth in monthly new editors (an increase of 2.1 new editors beyond the pre-integration baseline of 4.3), a 66% increase in monthly active editors (an increase of 6.3 from the pre-integration baseline of 9.5), and a 69% increase in number of new translated articles per editor² (an increase of 4.5 from the pre-integration baseline of 6.5). Therefore, the magnitude of these effects indicate that the increase in translation volume stems from both existing editors becoming more active and productive as well as an expansion of the editor base. This shift towards greater editor participation and productivity underscores Google Translate’s role in empowering Wikipedia editors.

Moreover, [Figure A.9](#) presents the temporal trend of editor productivity, providing a model-free and intuitive visual substantiation to our econometric estimates. The figure

¹For privacy consideration, only editor ID information is accessible on the website.

²Model (3) is analyzed at the editor level, as opposed to target language, and hence has a different number of observations

Table A.11: Editor Engagement and Productivity Analysis Post-Google Translate Integration

Dependent Variables: Model:	New Editor (1)	Active Editor (2)	Article/Editor (3)
GT Integration	2.1*** (0.6)	6.3*** (1.5)	4.5*** (1.3)
<i>Fixed-effects</i>			
Language	Yes	Yes	Yes
Year-Month	Yes	Yes	Yes
<i>Fit statistics</i>			
Observations	6,336	6,336	3,979
R ²	0.83264	0.91247	0.13120
Pre-integration Mean	4.3	9.5	6.5
<i>Clustered (Language) standard-errors in parentheses</i>			
<i>Signif. Codes: ***: 0.01, **: 0.05, *: 0.1</i>			

illustrates a steady temporal pattern in the monthly number of new translated articles per editor leading up to January 2019, after which there is a discernible upward trend. Individual editor productivity increases from an average of 6.8 article per editor/month in the pre-integration period to around 12 articles toward the end of our sample period. The error bars suggests variability in the data, yet the fitted LOESS curve shows that, on average, editors have been producing more articles each month since the integration of Google Translate.

Interestingly, the Interrupted Time Series analysis reveals a modest 10% increase in the level at the time of interruption for editor productivity ($\beta_2 = 0.7$ more articles relative to a 6.8 article pre-interruption baseline). However, there is a substantive positive shift between the post-interruption slope and pre-interruption slope ($\beta_3 = 0.19$ articles per editor/month). These findings imply that the full extent of the enhancement unfolds gradually as users become more adept with using the technology. Hence, while the tool lowers the skill barrier for editors to contribute, it should be complemented with strategic efforts to improve editor skill and experience with the new technology.

