

Uplifting News Ground Truth Dataset v1

Final Report & Validation

Report Generated: 2025-11-03 16:06:52

Executive Summary

- Total Articles: 7,715
- Unique Domains: 216
- Data Quality: Cleaned & Verified
- Labeler: Google Gemini 2.0 Flash
- Filter Version: filters/uplifting/v1

Quality Assurance

Check	Status
Blocked domains removed	PASS
Prompt leakage detected	PASS (0 articles)
Duplicates removed	PASS
Source diversity	PASS (216 domains)
Content length enforced	PASS (>=75 chars)

Tier Distribution

Tier	Count	Percentage
connection	5,133	66.5%
not_uplifting	2,528	32.8%
impact	54	0.7%

Top 20 Sources

Rank	Domain	Articles	%
1	reddit.com	1,415	18.3%
2	nos.nl	589	7.6%
3	dev.to	558	7.2%
4	ad.nl	526	6.8%
5	engadget.com	398	5.2%
6	phys.org	317	4.1%
7	fastcompany.com	313	4.1%
8	theverge.com	244	3.2%

Rank	Domain	Articles	%
9	rtl.nl	210	2.7%
10	upworthy.com	197	2.6%
11	cleantechica.com	192	2.5%
12	sciencedaily.com	176	2.3%
13	nature.com	137	1.8%
14	aws.amazon.com	125	1.6%
15	thebetterindia.com	119	1.5%
16	goodnewsnetwork.org	99	1.3%
17	genengnews.com	98	1.3%
18	insideclimateneWS.org	86	1.1%
19	electrek.co	77	1.0%
20	marginalrevolution.com	73	0.9%

Top 50 Impact Stories (Score 8.0)

These are the highest-scoring articles in the dataset. Review these to validate the quality and relevance of the training data.

1. World's First Archer Without All 4 Limbs: Payal Nag

Score: 8.28 | **Tier:** impact | **Source:** thebetterindia.com

URL: <https://thebetterindia.com/videos/lifestyle/worlds-first-archer-without-all-4-limbs-payal-nag-10599287>

Dimension Scores: - Agency: 10/10 - Resilience: 10/10 - Progress: 9/10 - Collective Benefit: 9/10 - Wonder: 8/10 - Innovation: 7/10 - Justice: 7/10 - Connection: 6/10

Key Strengths:

2. K8E Online....

Score: 8.24 | **Tier:** impact | **Source:** reddit.com

URL: https://www.reddit.com/r/programming/comments/1o80sif/k8e_online/

Dimension Scores: - Agency: 9/10 - Collective Benefit: 9/10 - Justice: 9/10 - Progress: 8/10 - Innovation: 8/10 - Resilience: 8/10 - Connection: 7/10 - Wonder: 7/10

Key Strengths:

3. Women Changemaker Of The Year : Kalpana Sankar

Score: 8.20 | **Tier:** impact | **Source:** thebetterindia.com

URL: <https://thebetterindia.com/videos/society/women-changemaker-of-the-year-kalpana-sankar-10518272>

Dimension Scores: - Agency: 10/10 - Collective Benefit: 10/10 - Progress: 9/10 - Connection: 8/10 - Resilience: 8/10 - Innovation: 7/10 - Justice: 7/10 - Wonder: 6/10

Key Strengths:

4. Radio Triyuga Revitalizes Indigenous Languages and Empowers Communities in Nepal

Score: 8.11 | **Tier:** impact | **Source:** culturalsurvival.org

URL: <https://www.culturalsurvival.org/news/radio-triyuga-revitalizes-indigenous-languages-and-empowers-communities-nepal>

Dimension Scores: - Agency: 9/10 - Collective Benefit: 9/10 - Resilience: 9/10 - Progress: 8/10 - Connection: 8/10 - Wonder: 8/10 - Innovation: 7/10 - Justice: 7/10

Key Strengths:

5. Fiji becomes the 26th country to eliminate trachoma as a public health problem

Score: 8.08 | **Tier:** impact | **Source:** who.int

URL: <https://www.who.int/news/item/20-10-2025-fiji-becomes-the-26th-country-to-eliminate-trachoma-as-a-public-health-problem>

Dimension Scores: - Collective Benefit: 10/10 - Agency: 9/10 - Progress: 9/10 - Connection: 8/10 - Resilience: 8/10 - Innovation: 7/10 - Justice: 7/10 - Wonder: 6/10

Key Strengths:

6. How an IAS Officer Transformed Lucknow into India's 3rd Cleanest City in Just

Score: 8.05 | **Tier:** impact | **Source:** thebetterindia.com

URL: <https://thebetterindia.com/civil-service/civil-service-heroes/lucknow-cleanliness-transformation-ias-indrajeet-singh-swachh-bharat-10504776>

Dimension Scores: - Agency: 9/10 - Progress: 9/10 - Collective Benefit: 9/10 - Connection: 8/10 - Innovation: 8/10 - Resilience: 8/10 - Justice: 7/10 - Wonder: 6/10

Key Strengths:

7. They thought their son lived a lonely life. But his funeral drew friends from ar

Score: 8.03 | **Tier:** impact | **Source:** upworthy.com

URL: <https://www.upworthy.com/disabled-son-lived-rich-inner-life-ex1>

Dimension Scores: - Connection: 10/10 - Agency: 9/10 - Collective Benefit: 9/10 - Progress: 8/10 - Resilience: 8/10 - Wonder: 8/10 - Innovation: 7/10 - Justice: 6/10

Key Strengths:

Source Diversity Analysis

Diversity in Top-N Articles

Metric	Value
Unique sources in top-10	6
Unique sources in top-20	8
Unique sources in top-50	12
Unique sources in full dataset	216

Top source in top-10: thebetterindia.com (4/10 = 40%)

Score Distribution Analysis

Score Range	Articles	%
9.0+	0	0.0%
8.0-8.9	7	0.1%
7.0-7.9	95	1.2%
6.0-6.9	968	12.5%
5.0-5.9	2,170	28.1%
Below 5.0	4,475	58.0%

Recommendations for Fine-Tuning

Dataset Suitability

READY FOR TRAINING

- High-quality articles: 54 impact-tier + 5,133 connection-tier = 5,187 total
- Source diversity: 216 unique domains (excellent)
- Data quality: All checks passed
- No contamination detected

Training Recommendations

1. **Positive Examples:** Use impact-tier (score > 8.0) articles as strong positives
 2. **Borderline Examples:** Use connection-tier (5.0 < score < 8.0) for nuanced training
 3. **Negative Examples:** Use not_uplifting tier (score < 5.0) as negatives
 4. **Validation Split:** Recommend 80/10/10 train/val/test split
 5. **Balance:** Consider class balancing based on tier distribution
-

Dataset Location

File: datasets/uplifting_ground_truth_v1/labeled_articles.jsonl

Size: 41 MB

Format: JSONL (one article per line)

Report generated by llm-distillery ground truth pipeline