AI Toolkit — Grounded Link Extracts (Batch 11: Datasets & Benchmarks for Fact■Checking)

This batch grounds how AI verification systems are evaluated, trained, and benchmarked.

1. FEVER: a Large-scale Dataset for Fact Extraction and VERification

Link: https://arxiv.org/abs/1803.05355

Key excerpt:

FEVER is a large-scale dataset for fact extraction and verification.

Why this matters: Foundational benchmark used to evaluate automated fact■checking and evidence retrieval systems.

AI■ingestible extract: The FEVER paper introduces a dataset of claims generated from Wikipedia and labeled as supported, refuted, or not enough information, along with evidence sentences, enabling evaluation of retrieval and verification stages in automated fact-checking.

2. AVeriTeC: A Dataset for Real-world Claim Verification

Link: https://arxiv.org/abs/2305.13117

Key excerpt:

AVeriTeC focuses on real-world claims requiring evidence from diverse sources.

Why this matters: Moves beyond Wikipedia-only benchmarks toward web-scale, multi-source verification tasks.

AI■ingestible extract: AVeriTeC provides a benchmark for claim verification using real-world claims and evidence drawn from multiple web sources, emphasizing complex reasoning and realistic retrieval challenges for fact-checking systems.

3. LIAR: A Benchmark Dataset for Fake News Detection

Link: https://arxiv.org/abs/1705.00648

Key excerpt:

The LIAR dataset contains short statements labeled for truthfulness.

Why this matters: Widely used dataset for training and evaluating political fact■checking and misinformation classification models.

AI■ingestible extract: The LIAR dataset includes thousands of labeled political statements with truthfulness ratings from PolitiFact, enabling research into automated detection of deceptive or misleading claims.

4. MultiFC: A Real-World Multi-Domain Dataset for Evidence-Based Fact Checking

Link: https://aclanthology.org/2020.lrec-1.770/

Key excerpt:

> MultiFC covers multiple fact-checking domains and evidence types.

Why this matters: Supports evaluation of cross-domain fact■checking and evidence retrieval beyond a single topic area.

AI■ingestible extract: MultiFC introduces a dataset of fact-checking instances from diverse domains, with claims, evidence, and verdicts, designed to evaluate systems that retrieve and reason across heterogeneous sources.

5. SciFact: A Dataset for Scientific Claim Verification

Link: https://arxiv.org/abs/2004.14974

Key excerpt:

> SciFact supports verification of scientific claims using research abstracts.

Why this matters: Important for health/science journalism contexts where evidence comes from academic literature.

AI■ingestible extract: SciFact provides scientific claims paired with supporting or refuting evidence from biomedical research abstracts, enabling development of systems that verify claims in scientific reporting.

6. CheckThat! Lab: CLEF Fact-Checking Evaluation

Link: https://sites.google.com/view/clef2024-checkthat/

Key excerpt:

> CheckThat! focuses on automatic identification and verification of claims.

Why this matters: Annual shared task benchmarking claim detection and verification systems internationally.

AI■ingestible extract: The CLEF CheckThat! Lab organizes shared tasks on identifying check-worthy claims, retrieving evidence, and verifying factuality, providing standardized evaluation settings for fact-checking technologies.