Solstice: End-to-End Pipeline for Medical Document Fact-Checking

From PDF Ingestion to Multi-Agent Verification

1 Introduction

Solstice is an AI-native system that transforms unstructured medical PDFs into fact-checked, evidence-backed documents. The pipeline combines computer vision for layout understanding, natural language processing for text extraction, and orchestrated multi-agent systems for claim verification.

2 Stage 1: PDF Document Ingestion

2.1 Layout Detection with Detectron2

The pipeline begins with sophisticated layout analysis using Detectron2, a state-of-the-art object detection framework. Below is an actual example from a scientific paper processed through our pipeline:

The detection process uses:

- Faster R-CNN with ResNet-50 backbone
- IoU threshold of 0.7 for overlap resolution
- Hierarchical nesting for complex layouts
- Custom post-processing for medical documents

2.2 Text and Visual Extraction

After layout detection, the pipeline extracts content:

1. **Text Extraction**: PyMuPDF extracts text within bounding boxes
2. **OCR Correction**: SymSpell fixes common OCR errors (0→O, l→I)
3. **Figure/Table Export**: Visual elements saved as 300 DPI PNGs
4. **Reading Order**: Column detection determines logical flow

3 Stage 2: Multi-Agent Fact-Checking

3.1 Agent Architecture

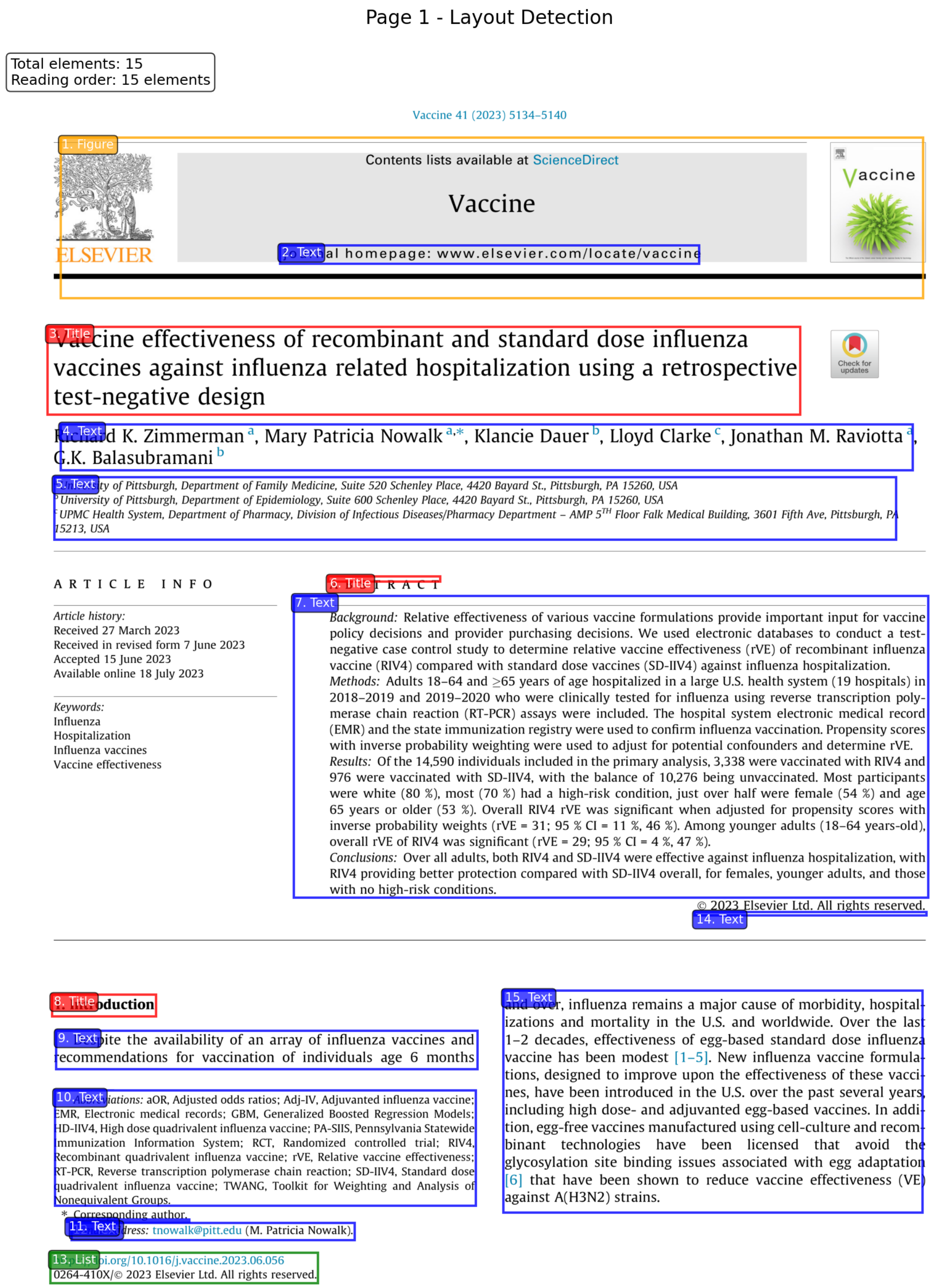
The fact-checking system employs specialized agents in an orchestrated pipeline:

3.2 Agent Responsibilities

3.2.1 Evidence Extractor

```
# Searches for claim-relevant quotes
async def extract_evidence(claim, document):
    model = "gpt-4"
    temperature = 0  # Deterministic

    quotes = search_document(claim, document)
    return preserve_exact_quotes(quotes)
```


Page 1 - Layout Detection

Total elements: 15
Reading order: 15 elements

Vaccine 41 (2023) 5134–5140

1. Figure

Contents lists available at ScienceDirect

Vaccine

2. Text al homepage: www.elsevier.com/locate/vaccine

ELSEVIER

3. Title ccine effectiveness of recombinant and standard dose influenza vaccines against influenza related hospitalization using a retrospective test-negative design

4. Text d K. Zimmerman [a], Mary Patricia Nowalk [a,*], Klancie Dauer [b], Lloyd Clarke [c], Jonathan M. Raviotta [a], .K. Balasubramani [b]

5. Text ty of Pittsburgh, Department of Family Medicine, Suite 520 Schenley Place, 4420 Bayard St., Pittsburgh, PA 15260, USA
University of Pittsburgh, Department of Epidemiology, Suite 600 Schenley Place, 4420 Bayard St., Pittsburgh, PA 15260, USA
UPMC Health System, Department of Pharmacy, Division of Infectious Diseases/Pharmacy Department – AMP 5TH Floor Falk Medical Building, 3601 Fifth Ave, Pittsburgh, PA 15213, USA

ARTICLE INFO

Article history:
Received 27 March 2023
Received in revised form 7 June 2023
Accepted 15 June 2023
Available online 18 July 2023

Keywords:
Influenza
Hospitalization
Influenza vaccines
Vaccine effectiveness

6. Title RACT

7. Text

Background: Relative effectiveness of various vaccine formulations provide important input for vaccine policy decisions and provider purchasing decisions. We used electronic databases to conduct a test-negative case control study to determine relative vaccine effectiveness (rVE) of recombinant influenza vaccine (RIV4) compared with standard dose vaccines (SD-IIV4) against influenza hospitalization.
Methods: Adults 18–64 and ≥65 years of age hospitalized in a large U.S. health system (19 hospitals) in 2018–2019 and 2019–2020 who were clinically tested for influenza using reverse transcription polymerase chain reaction (RT-PCR) assays were included. The hospital system electronic medical record (EMR) and the state immunization registry were used to confirm influenza vaccination. Propensity scores with inverse probability weighting were used to adjust for potential confounders and determine rVE.
Results: Of the 14,590 individuals included in the primary analysis, 3,338 were vaccinated with RIV4 and 976 were vaccinated with SD-IIV4, with the balance of 10,276 being unvaccinated. Most participants were white (80 %), most (70 %) had a high-risk condition, just over half were female (54 %) and age 65 years or older (53 %). Overall RIV4 rVE was significant when adjusted for propensity scores with inverse probability weights (rVE = 31; 95 % CI = 11 %, 46 %). Among younger adults (18–64 years-old), overall rVE of RIV4 was significant (rVE = 29; 95 % CI = 4 %, 47 %).
Conclusions: Over all adults, both RIV4 and SD-IIV4 were effective against influenza hospitalization, with RIV4 providing better protection compared with SD-IIV4 overall, for females, younger adults, and those with no high-risk conditions.

© 2023 Elsevier Ltd. All rights reserved.
14. Text

8. Title oduction

9. Text ite the availability of an array of influenza vaccines and ecommendations for vaccination of individuals age 6 months

10. Text iations: aOR, Adjusted odds ratios; Adj-IV, Adjuvanted influenza vaccine; EMR, Electronic medical records; GBM, Generalized Boosted Regression Models; HD-IIV4, High dose quadrivalent influenza vaccine; PA-SIIS, Pennsylvania Statewide Immunization Information System; RCT, Randomized controlled trial; RIV4, Recombinant quadrivalent influenza vaccine; rVE, Relative vaccine effectiveness; RT-PCR, Reverse transcription polymerase chain reaction; SD-IIV4, Standard dose quadrivalent influenza vaccine; TWANG, Toolkit for Weighting and Analysis of Nonequivalent Groups.
* Corresponding author.
11. Text dress: tnowalk@pitt.edu (M. Patricia Nowalk).

13. List i.org/10.1016/j.vaccine.2023.06.056
0264-410X/© 2023 Elsevier Ltd. All rights reserved.

15. Text r, influenza remains a major cause of morbidity, hospitalizations and mortality in the U.S. and worldwide. Over the last 1–2 decades, effectiveness of egg-based standard dose influenza vaccine has been modest [1–5]. New influenza vaccine formulations, designed to improve upon the effectiveness of these vaccines, have been introduced in the U.S. over the past several years, including high dose- and adjuvanted egg-based vaccines. In addition, egg-free vaccines manufactured using cell-culture and recombinant technologies have been licensed that avoid the glycosylation site binding issues associated with egg adaptation [6] that have been shown to reduce vaccine effectiveness (VE) against A(H3N2) strains.

Figure 1: Real layout detection output from Zimmerman et al. (2023) showing detected text blocks, tables, and figures with bounding boxes and confidence scores

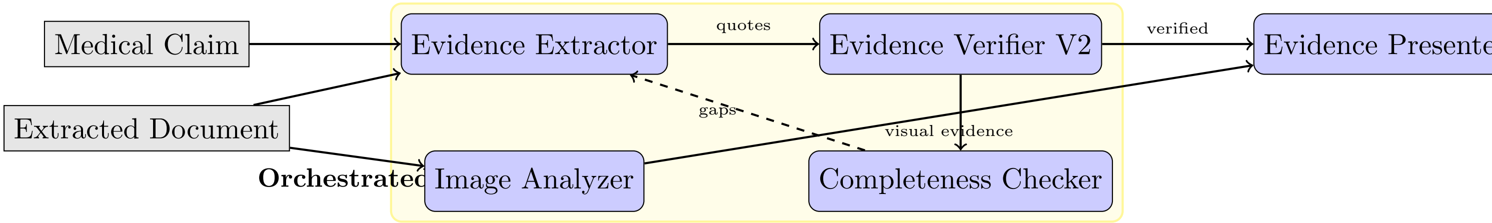


Figure 2: Multi-agent system architecture with feedback loops for comprehensive evidence extraction

3.2.2 Evidence Verifier V2

Validates that extracted quotes exist in the source:

- Semantic matching for OCR variations
- Filters tangential content
- Returns verification statistics

3.2.3 Completeness Checker

Identifies gaps and triggers additional extraction:

- Reviews verified evidence coverage
- Searches for missing aspects
- Feeds findings back to pipeline

3.2.4 Image Evidence Analyzer

Processes visual elements with vision models:

```
# Analyzes figures and tables
async def analyze_image(image_path, claim):
    model = "claude-3"  # Multimodal

    analysis = await model.analyze(
        image=load_image(image_path),
        claim=claim
    )
    return {
        "supports_claim": analysis.relevant,
        "explanation": analysis.details
    }
```

4 Stage 3: Output Generation

4.1 Evidence Presentation

The final stage consolidates all evidence into structured outputs:

1. **JSON Output**: Machine-readable evidence with metadata
2. **HTML Reports**: Human-readable with embedded images
3. **Coverage Assessment**: Complete/partial/none ratings
4. **Confidence Scores**: Based on evidence quantity/quality

4.2 Marketing Material Processing

The system includes a specialized pipeline for marketing materials with enhanced visual element detection:

Marketing pipeline differences:

- Separate cache directory (data/marketing_cache/)

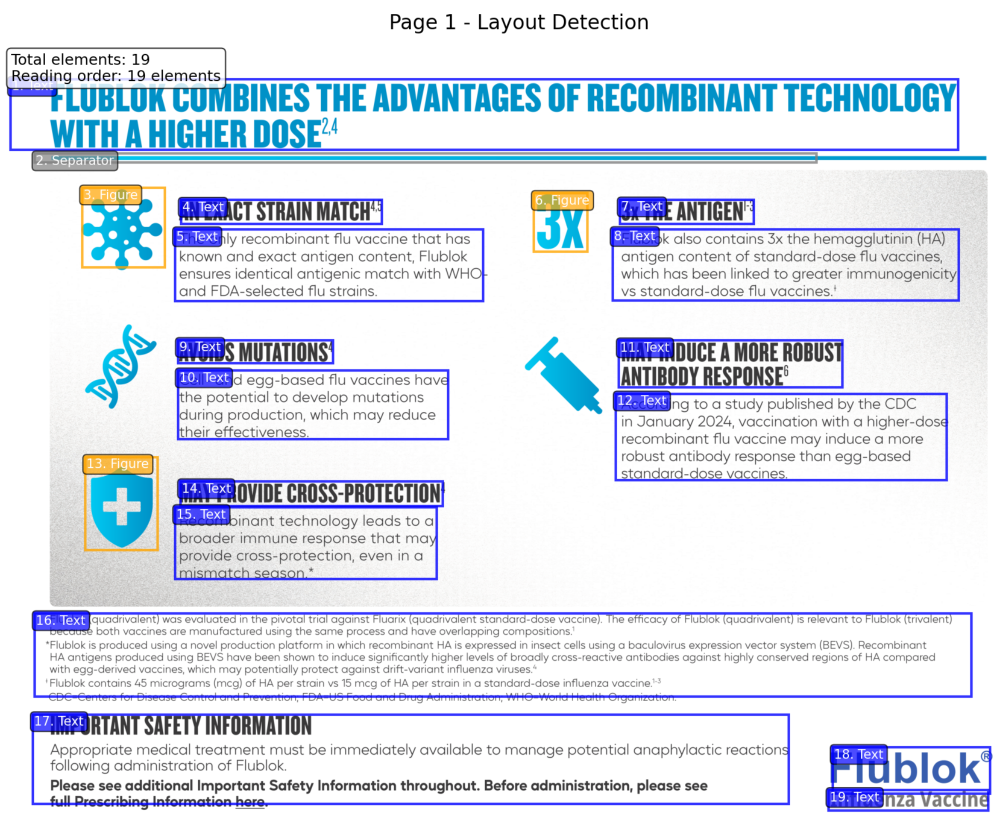


Figure 3: Marketing material layout detection from Flublok one-pager showing detection of promotional graphics, key benefit callouts, and visual hierarchy elements

- Enhanced visual element extraction
- Claim detection from promotional language
- Cross-referencing with clinical evidence

5 Implementation Details

5.1 Orchestration and Parallelism

```python
# Parallel claim processing with semaphore
async def process_claims(claims, documents):
    semaphore = asyncio.Semaphore(2)

    async def process_single(claim):
        async with semaphore:
            return await claim_orchestrator.run(
                claim, documents
            )

    results = await asyncio.gather(*[
        process_single(c) for c in claims
    ])
    return results
```

5.2 Error Handling and Reliability

- **Retry Logic**: Exponential backoff for API failures
- **JSON Parsing**: Handles markdown-wrapped responses
- **Token Management**: Dynamic adjustment for long documents
- **Cache System**: All intermediate results persisted

6 Real-World Applications

Current deployments include:

1. **Vaccine Studies**: Efficacy claims vs clinical trials
2. **Drug Safety**: Package inserts vs FDA documents
3. **Medical Devices**: Marketing claims vs regulatory filings
4. **Treatment Guidelines**: Recommendations vs evidence base

7 Performance Characteristics

- **Layout Detection**: 95%+ mAP on medical PDFs
- **Text Extraction**: 99%+ accuracy with OCR correction
- **Evidence Verification**: 80-90% verification rates
- **Processing Speed**: 2-3x speedup with parallelization
- **Image Analysis**: 30-40% of images contain relevant evidence

8 Conclusion

Solstice demonstrates how modern AI capabilities can be orchestrated to solve complex document processing challenges. By combining layout understanding, multimodal analysis, and agent-based verification, the system provides reliable fact-checking for medical documentation while maintaining full traceability through comprehensive caching.