Mira 2.0 WBP Assertion Framework

Announcement & Summary Report

Author: Chin-Yew Lin
Date: November 29, 2025

Executive Summary

We have successfully migrated **2,318 assertions** across **224 meetings** from Kening's original format to Chin-Yew's standardized **WBP (Workback Plan) Evaluation Framework**, achieving a **99.6% GPT-5 conversion rate** and **99.3% Phase 1 alignment**.

---

1. WBP Assertion Framework Migration

Conversion Results

Metric	Value
Total Meetings	224
Total Assertions	2,318
GPT-5 Conversion Rate	**99.6%**
Phase 1 Alignment	**99.3%**
Original Dimensions	232 (fragmented)
Standardized Dimensions	12 (consolidated)

Key Improvements

Problem (Kening's Original)	Solution (WBP Framework)
232 inconsistent dimension names	14 well-defined dimensions with clear codes (S1-S19, G1-G5)
Hardcoded dates and names	Template-based reusable assertions
Only ~3% grounding checks	Dedicated Grounding layer (G1-G5)
Inconsistent phrasing	Standardized evaluation criteria

Dimension Distribution

Dimension	Name	Count	%
S2	Timeline Alignment	449	19.4%
S3	Ownership Assignment	414	17.9%
S4	Deliverables & Artifacts	399	17.2%
S6	Dependencies & Blockers	322	13.9%

S1	Meeting Details	321	13.8%
S19	Caveat & Clarification	279	12.0%
G5	Hallucination Check	60	2.6%
S11	Risk Mitigation Strategy	42	1.8%

Layer Distribution:

- Structural (S): 97.2% (2,252 assertions)
- Grounding (G): 2.8% (66 assertions)

---

2. Mira 2.0 Visualization Tool

New Streamlit Application (`mira2.py`)

A dedicated viewer for WBP assertions with the following features:

- **Color-coded dimension badges**
  - 🔵 **Blue** = Structural dimensions (S1-S19)
  - 🟣 **Purple** = Grounding dimensions (G1-G5)
- **Smart assertion sorting**
  - Structural assertions (S) displayed before Grounding (G)
  - Sub-indices for same-dimension assertions (S4-1, S4-2, S4-3)
- **Duplicate navigation**
  - `[1/N]` markers in sidebar to navigate response variations
  - Easy comparison across different LLM outputs for same query

Quick Start

```
streamlit run mira2.py
```

---

3. Dataset Analysis Findings

Near-Duplicate Structure

Metric	Value
Total Entries	224
Unique User Queries	47
Duplicate Groups	44
Avg Response Variations	~4.8 per query
Unique Users	17

Key Finding

Duplicates are intentional design, not errors:

- Same user utterance (query)
- Different LLM responses
- Enables comparative evaluation of assertion consistency across response variations

This structure supports:

- ✓ Comparative analysis of assertion quality
- ✓ Evaluation of LLM consistency
- ✓ Rich annotation opportunities

---

4. Key Reports & Files

🏷 Documentation

Document	Location
Migration Report (Main)	`docs/ChinYew/Assertion_Framework_Migration_Report.md`
Data Analysis Report	`docs/ChinYew/DATA_ANALYSIS_REPORT.md`

🖥 Application

File	Description
`mira2.py`	WBP Assertion Viewer (Streamlit)

📦 Data Files

File	Description
`docs/ChinYew/assertions_converted_full.jsonl`	224 meetings with WBP-mapped assertions
`docs/ChinYew/Assertions_genv2_for_LOD1126part1.jsonl`	Kening's original assertions

---

5. Recommendations

1. **Use Phase 1 Framework** - 99.3% of assertions already map to Phase 1 dimensions
2. **Consider consolidating S5 into S2** - The 16 S5 (Task Dates) assertions overlap with S2 (Timeline Alignment)
3. **Add more Grounding assertions** - Currently only 2.8% are grounding-layer; consider generating more G1-G4 assertions

---

Appendix: Dimension Reference

Structural Dimensions (S)

Code	Name
S1	Meeting Details
S2	Timeline Alignment
S3	Ownership Assignment
S4	Deliverables & Artifacts
S5	Task Dates
S6	Dependencies & Blockers
S11	Risk Mitigation Strategy
S18	Post-Event Actions
S19	Caveat & Clarification

Grounding Dimensions (G)

Code	Name
G1	Attendee Grounding
G4	Topic Grounding
G5	Hallucination Check

For detailed technical information, see the full Migration Report at `docs/ChinYew/Assertion_Framework_Migration_Report.md`