

# Assignment 3 – Data Pipeline Deliverable

**Project: Madison Agent: Model Behavior Evaluation System**

**Problem: API teams lack integrated tooling to evaluate LLM behavior consistently against real-world prompts and high-quality reference responses.**

## Overview

This project builds a reproducible n8n data pipeline that collects, cleans, validates, and stores datasets required by the Madison Agent. The goal is to provide:

1. Real-world, continuously updated AI content (research + news)
2. Ground-truth instruction–response examples for behavioral comparison

The workflow runs locally in n8n, integrates three Tier-1 data sources, applies validation and normalization, and exports the results in structured CSV and JSON formats for reuse in Assignment 4.

Data is collected once and saved to disk to avoid repeated API calls and rate-limit issues.

## Part 1 – Working n8n Data Collection Workflow

### Data Sources

#### Source 1: ArXiv Computer Science – AI RSS Feed

- Type: Academic research papers (title, abstract, URL, publication date)
- Amount: 100 records
- Purpose: Provides current research context and emerging model capabilities
- Collection Method: RSS feed via n8n HTTP Request + XML node
- Quality: 100/100 complete records

## **Source 2: Smol AI News RSS Feed**

- Type: Industry AI news and research updates
- Amount: 10 records
- Purpose: Adds practical industry context and deployment trends
- Collection Method: RSS feed via n8n HTTP Request + XML node
- Quality: 10/10 complete records

## **Source 3: AlpacaEval Ground Truth Dataset (HuggingFace)**

- Type: Instruction–response pairs with evaluation metadata
- Amount: 80 records
- Purpose: Provides high-quality baseline examples for evaluating model outputs
- Collection Method: HTTP Request to HuggingFace public dataset API + Code nodes
- Quality: 80/80 complete records

## **Workflow Summary**

The n8n workflow performs the following:

1. Collects RSS content from ArXiv and Smol AI
2. Fetches AlpacaEval examples from HuggingFace
3. Normalizes schemas across RSS feeds
4. Extracts and flattens AlpacaEval records
5. Applies validation rules:
  - Empty summaries removed
  - Empty Alpaca responses removed
  - Minimum length checks enforced
  - Duplicates removed by URL
  - Dates standardized
6. Outputs:
  - RSS content → CSV
  - AlpacaEval → JSON

The workflow runs locally and is reproducible.

---

# Part 2 – Data Documentation

## Final Schemas ([Madison Agent Data Inventory Markdown](#))

### Normalized Content Feed (CSV)

Fields:

```
source  
title  
url  
published_at  
summary  
raw
```

Notes:

- `summary` is unified from RSS description/content fields
- `raw` contains the full original provider payload stored as a single column
- Duplicate URLs removed
- Records without summaries dropped
- Dates standardized

### AlpacaEval Dataset (JSON)

Fields:

```
record_id  
record_type  
source  
prompt  
response  
model_used  
category  
dataset_origin  
response_length  
quality_score  
instruction_following_score  
coherence_score  
helpfulness_score
```

`collected_at`  
`quality_flag`

**Notes:**

- Nested structures flattened
- Empty responses removed
- Prompts < 20 chars dropped
- Responses < 50 chars dropped
- Category inferred heuristically
- Scores included as provided placeholders
- Dataset diversity enforced

## Data Validation & Cleaning

### RSS Feeds

- Removed rows with empty summaries
- Deduplicated by URL
- Unified summary fields
- Standardized publication dates
- Preserved raw provider payload in single column
- Filtered for relevance
- Final count after validation: 110 records

### AlpacaEval

- Removed records with empty prompts or responses
- Enforced minimum prompt length (20 chars)
- Enforced minimum response length (50 chars)
- Flattened nested structures
- Ensured dataset diversity
- Final count: 80 records

## Statistics

- Total records: 190
- Clean records: 190 (100% quality rate)
- Sources: 3
- Output formats: CSV + JSON

# How This Dataset Supports Madison Agents

The Madison Agent evaluates LLM behavior using two complementary data types:

## Context Data (RSS Feeds)

ArXiv and Smol AI provide continuously updated, real-world prompts and topics representing:

- Current research trends
  - Deployment challenges
  - Emerging model capabilities
- These feeds supply realistic evaluation scenarios.

## Ground Truth Data (AlpacaEval)

AlpacaEval provides high-quality instruction–response examples that act as behavioral baselines. Madison compares live model outputs against these reference responses to assess:

- Instruction following
- Coherence
- Helpfulness
- Overall quality

Together, these datasets allow Madison to evaluate models against both real-world content and curated ground truth.

## Storage Format

Data is stored separately due to structural differences:

- RSS content exported as CSV for easy inspection and filtering
- AlpacaEval exported as JSON to preserve evaluation metadata

Files produced:

- `content_feed.csv`
- `Alpaca_eval.json`

Both are reusable inputs for Assignment 4.

---

# Part 3 – Data Quality

## Quality Metrics

- Completeness: 100%
- Required fields present in all records
- Duplicate URLs removed
- Dates standardized
- Empty responses removed
- Minimum length enforced

## Validation Checks

- RSS summaries present
- Alpaca prompts/responses present
- Response length verified
- Dataset diversity maintained

# Setup Guide

## Requirements

Node.js installed on your system.

n8n installed locally by running `npm install -g n8n`, then starting it with `n8n start`.

## How to Run the Workflow

1. Launch n8n locally in your browser at <http://localhost:5678>.
2. Import Ravi\_Aravind\_A3\_Workflow.json into n8n.
3. Click “Execute Workflow” to start the pipeline.
4. Wait for all nodes to complete successfully.
5. Once execution finishes, the exported files will appear in the configured output directory:
  - content\_feed.csv
  - alpaca\_eval.json

*Tested locally on n8n v1.64.3 (installed via Homebrew)*

## Credentials

No API keys are required. All data sources used in this project are Tier-1 public RSS feeds and public HuggingFace datasets.

## **Common Errors & Fixes**

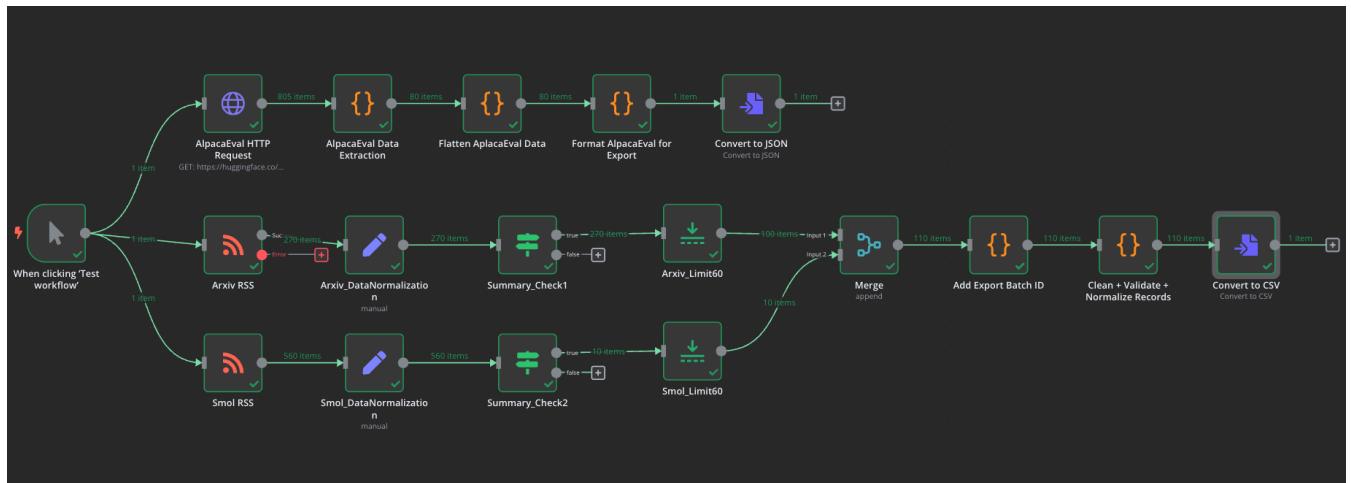
1. RSS rate limiting is avoided by running the workflow once and saving outputs locally instead of repeatedly re-fetching data.
2. XML parsing issues are handled using HTTP status checks and XML conversion nodes before normalization.
3. Null or short summaries are filtered out in the validation code node.
4. Duplicate URLs are removed during the normalization step to ensure each record is unique.

## **Error Handling**

The pipeline includes defensive validation to prevent malformed inputs from entering exports. HTTP status checks prevent XML parsing failures, summary validation nodes remove empty content, and URL-based deduplication avoids duplicate records. These safeguards ensure the workflow completes reliably even when upstream feeds change structure.

# Demo Documentation/ Walkthrough

## 1. Full workflow overview



*This screenshot shows the complete n8n workflow. Three parallel data pipelines ingest ArXiv research RSS, Smol AI news RSS, and AlpacaEval ground truth from HuggingFace. Each branch applies normalization, validation, and filtering before merging outputs and exporting structured CSV and JSON files. This architecture enables collecting real-world context and curated evaluation data in a single reproducible pipeline.*

## 2. RSS nodes

The screenshot shows the Arxiv RSS step configuration and its resulting output table.

**INPUT** (Left Panel): A dropdown menu titled "When clicking Test workflow" showing "1 item". Below it is a message: "This is an item, but it's empty."

**Arxiv RSS Step Configuration (Top Bar):** Includes tabs for "Parameters", "Settings", "Docs", and a red "Test step" button.

**OUTPUT (Right Panel):** Shows the results of the step execution.

**Success Branch (270 items):** The main table view.

**Error Branch:** No errors present.

**Table Headers:** creator, rights, title, link, pubDate, dc:creator

**Table Data (First Row):**

creator	rights	title	link	pubDate	dc:creator
Peter Zeng, Welling Li, Amie Paige, Zhengxiang Wang, Panagiotis Kallosis, Dimitris Samaras, Gregory Zelinsky, Susan Brennan, Owen Rambow	http://creativecommons.org/licenses/by/4.0 /LVLMs and Humans Ground Differently in Referential Communication	LVLMs and Humans Ground Differently in Referential Communication	https://arxiv.org/abs/2601.19792	Fri, 30 Jan 2026 00:00:00 -0500	Peter Zeng, Welling Li, Amie Paige, Zhengxiang Wang, Panagiotis Kallosis, Dimitris Samaras, Gregory Zelinsky, Susan Brennan, Owen Rambow

**Page Navigation:** 1 2 3 4 ... 27 > Page Size | 10 ⌂

**a) ArXiv RSS Read Node:** This node pulls the latest Computer Science AI research articles from ArXiv using an RSS feed. It retrieves titles, abstracts, URLs, and publication dates, forming the primary academic context dataset used for evaluating model knowledge of current research trends.

**INPUT**

When clicking "Test workflow" 1 item

**Smol RSS**

Parameters Settings Docs

URL: https://news.smol.ai/rss.xml

Options

No properties Add option

**OUTPUT**

560 items

title	link	pubDate	content:encoded	content:encodedSnippet
MoltBook takes over the timeline	https://news.smol.ai/issues/26-01-30-moltbook/	Fri, 30 Jan 2026 05:44:39 GMT	<p><strong>Moltbook takes over the timeline.</strong></p><div><blockquote><div><h1>AI News for 1/29/2026-1/30/2026</h1><div><ul><li>12 subreddits, <a href="https://twitter.com/i/lists/1585430245762441216"><strong>544</strong> channels, and <strong>7413</strong> messages for you. Estimated reading time saved (at 200wpm): <strong>657 minutes</strong>. <strong>Our new website</strong> is now up with full metadata search and beautiful vibe coded presentation of all past issues. See https://news.smol.ai for the full news breakdowns and give us feedback on <a href="https://x.com/Smol_AI">@smol_ai</a></div></div></div>	Moltbook takes over the timeline. News for 1/29/2026-1/30/2026. We checked 12 subreddits, <a href="https://twitter.com/i/lists/1585430245762441216"><strong>544</strong> channels, and <strong>7413</strong> messages for you. Estimated reading time saved (at 200wpm): <strong>657 minutes</strong>. <strong>Our new website</strong> is now up with full metadata search and beautiful vibe coded presentation of all past issues. See https://news.smol.ai for the full news breakdowns and give us feedback on <a href="https://x.com/Smol_AI">@smol_ai</a></div></div></div>

**b) Smol RSS Read Node:** This node ingests AI industry news from Smol AI's RSS feed. These records provide practical deployment context and real-world AI developments, complementing academic research from ArXiv.

**INPUT**

AlpacaEval HTTP Request 805 items

dataset	instruction	output	generator
helpful_base	What are the names of some famous actors that started their careers on Broadway?	Some famous actors that started their careers on Broadway include: \n1. Hugh Jackman \n2. Meryl Streep \n3. Denzel Washington \n4. Julia Roberts \n5. Christopher Walken \n6. Anthony Rapp \n7. Audra McDonald \n8. Nathan Lane \n9. Sarah Jessica Parker \n10. Lin-Manuel Miranda	text_davinci_003
helpful_base	How did US states get their names?	US states get their names from a variety of sources, including Native American tribes, Spanish explorers, British colonists, and even presidents. For example, the state of Alabama was named after the Native American tribe that lived in the area, while the state of Florida gets its name from the Spanish explorer, Ponce de Leon, who explored the area in the 1500s. Other states are named after English kings (like Virginia, named after England's "Virgin Queen," Queen Elizabeth I) or presidents (like Washington, named after George Washington).	text_davinci_003

**AlpacaEval Data Extraction**

Test step

Parameters Settings Docs

Mode: Run Once for All Items

Language: JavaScript

```

1 // Extract 80 evaluations from AlpacaEval
2 const items = $input.all();
3
4 console.log("== Starting AlpacaEval Extraction ==");
5 console.log(`Total items available: ${items.length}`);
6
7 const evaluations = [];
8 const categories = {};
9
10 // Process items
11 for (let i = 0; i < items.length && evaluations.length < 80; i++) {
12   const example = items[i].json;
13
14   // Get the fields
15   const instruction = example.instruction;
16   const output = example.output;
17   const generator = example.generator || 'text_davinci_003';
18   const datasetName = example.dataset || 'helpful_base';
19
20   // Advanced validation

```

Type \$ for a list of special vars/methods. Debug by using console.log() statements and viewing their output in the browser console.

**OUTPUT**

80 items

record_id	record_type	source	data
eval_alpaca_1	evaluation	alpaca_eval_huggingface	<p><b>prompt:</b> What are the names of some famous actors that started their careers on Broadway?</p> <p><b>response:</b> Some famous actors that started their careers on Broadway include: \n1. Hugh Jackman \n2. Meryl Streep \n3. Denzel Washington \n4. Julia Roberts \n5. Christopher Walken \n6. Anthony Rapp \n7. Audra McDonald \n8. Nathan Lane \n9. Sarah Jessica Parker \n10. Lin-Manuel Miranda</p> <p><b>model_used:</b> text_davinci_003</p> <p><b>scores:</b></p> <ul style="list-style-type: none"> <li><b>quality:</b> 0.85</li> <li><b>instruction_following:</b> 0.9</li> <li><b>coherence:</b> 0.88</li> <li><b>helpfulness:</b> 0.87</li> </ul> <p><b>category:</b> explanation</p> <p><b>dataset_origin:</b> helpful_base</p> <p><b>response_length:</b> 264</p>
eval_alpaca_2	evaluation	alpaca_eval_huggingface	<p><b>prompt:</b> How did US states get their names?</p> <p><b>response:</b> US states get their names from a variety of sources, including Native American tribes, Spanish explorers, British colonists, and even presidents. For example, the state of Alabama was named after the Native American tribe that lived in the area, while the state of Florida gets its name from the Spanish explorer, Ponce de Leon, who explored the area in the 1500s. Other states are named after English kings (like Virginia, named after England's "Virgin Queen," Queen Elizabeth I) or presidents (like Washington, named after George Washington).</p>

**c) AlpacaEval Extraction Node:** This Code node processes raw AlpacaEval records from HuggingFace. It filters incomplete examples, enforces minimum prompt/response lengths, categorizes instructions, and selects a diverse subset of 80 evaluation pairs. These examples act as ground truth references for later model comparison.

### 3. Cleaning nodes

The screenshot shows the AlpacaEval Data Extraction interface with the "Flatten AlpacaEval Data" node selected. The input dataset contains two rows of AlpacaEval data, each with fields like record\_id, record\_type, source, prompt, response, etc. The output dataset shows the flattened version of this data, where nested structures are removed and placed at the top level. The code editor in the center contains a JavaScript script that performs this flattening.

record_id	record_type	source	prompt	response
eval_alpaca_1	evaluation	alpaca_eval_huggingface	What are the names of some famous actors that started their careers on Broadway?	Some famous actors that started their careers on Broadway include: \n1. Hugh Jackman \n2. Meryl Streep \n3. Denzel Washington \n4. Julia Roberts \n5. Christopher Walken \n6. Anthony Rapp \n7. Audra McDonald \n8. Nathan Lane \n9. Sarah Jessica Parker \n10. Lin-Manuel Miranda
eval_alpaca_2	evaluation	alpaca_eval_huggingface	How did US states get their names?	US states get their names from a variety of sources, including Native tribes, Spanish and British colonists, presidents, and state of Alabama, which was named after the American tribe that lived in the area, while Florida gets its name from the Spanish explorer de Leon, who explored the area in the 1500s. Some states are named after English kings (like England) and some are named after English queens (like Queen Victoria).

**a) Flatten AlpacaEval Data Node:** This node flattens nested AlpacaEval structures into a tabular schema, bringing prompts, responses, scores, and metadata to the top level. This makes the dataset easier to export, validate, and reuse in downstream evaluation agents.

The screenshot shows the AlpacaEval Data Extraction interface with the "Format AlpacaEval for Export" node selected. The input dataset is the same as in the previous screenshot. The output dataset is a JSON object with "collection\_metadata" and "records" fields. The "collection\_metadata" field contains general information about the dataset, and the "records" field contains the flattened AlpacaEval data from the input.

record_id	record_type	source	prompt	response
eval_alpaca_1	evaluation	alpaca_eval_huggingface	What are the names of some famous actors that started their careers on Broadway?	Some famous actors that started their careers on Broadway include: \n1. Hugh Jackman \n2. Meryl Streep \n3. Denzel Washington \n4. Julia Roberts \n5. Christopher Walken \n6. Anthony Rapp \n7. Audra McDonald \n8. Nathan Lane \n9. Sarah Jessica Parker \n10. Lin-Manuel Miranda
eval_alpaca_2	evaluation	alpaca_eval_huggingface	How did US states get their names?	US states get their names from a variety of sources, including Native tribes, Spanish and British colonists, presidents, and state of Alabama, which was named after the American tribe that lived in the area, while Florida gets its name from the Spanish explorer de Leon, who explored the area in the 1500s. Some states are named after English kings (like England) and some are named after English queens (like Queen Victoria).

**b) Format Alpaca Eval for Export Node:** This node packages AlpacaEval records into a final JSON structure with collection metadata, quality statistics, schema definitions, and validation checks. The output becomes a reusable ground truth evaluation dataset.

**Arxiv\_DataNormalization**

**INPUT**

**OUTPUT**

**Parameters**

- source: Arxiv
- title: {{ \$json.title }}
- url: {{ \$json.link }}
- published\_at: {{ \$json.publishDate }}
- summary: {{ \$json.content }}
- raw: {{ \$json }}

**Output Fields**

source	title	url	published_at	summary	raw
Arxiv	LVLMs and Humans Ground Differently in Referential Communication	https://arxiv.org/abs/2601.19792	Fri, 30 Jan 2026 00:00:00 -0500	arXiv:2601.19792v2 Announce Type: replace-cross \nAbstract: For generative AI agents to partner effectively with human users, the ability to accurately predict human intent is critical. But this ability to collaborate remains limited by a critical deficit: an inability to model common ground. Here, we present a referential communication experiment with a factorial design involving director-matcher pairs (human-human, human-AI, AI-human, and AI-AI) that interact with multiple turns in repeated rounds to match pictures of objects not associated with any obvious lexicalized labels. We release the online pipeline for data collection, the tools and analyses for accuracy, efficiency, and lexical overlap, and a corpus of 356 dialogues (89 pairs over 4 rounds each) that unmasks LVLMs' limitations in interactively resolving referring expressions, a crucial skill that	creator : Peter Zeng, Welling Li, Amie Paige, Zhengxiang Wang, Panagiotis Kalilosis, Dimitris Samaras, Gregory Zelinsky, Susan Brennan, Owen Rambow rights : http://creativecommons.org/licenses/by/4.0/ title : LVLMs and Humans Ground Differently in Referential Communication link : https://arxiv.org/abs/2601.19792 pubdate : Fri, 30 Jan 2026 00:00:00 -0500 dc.creator : Peter Zeng, Zhengxiang Wang, Dimitris Samaras, Susan Brennan, Owen Rambow content : arXiv:2601.19792v2 Announce Type: replace-cross \nAbstract: For generative AI agents to partner effectively with human users, the ability to accurately predict human intent is critical. But this ability to collaborate remains limited by a critical deficit: an inability to model common ground. Here, we present a referential communication experiment with a factorial design involving director-matcher pairs (human-human, human-AI, AI-human, and AI-AI) that interact with multiple turns in repeated rounds to match pictures of objects not associated with any obvious lexicalized labels. We release the online pipeline for data collection, the tools and analyses for accuracy, efficiency, and lexical overlap, and a corpus of 356 dialogues (89 pairs over 4 rounds each) that unmasks LVLMs' limitations in interactively resolving referring expressions, a crucial skill that

**c) ArXiv / Smol Data Normalization Nodes:** These nodes normalize RSS fields into a unified schema (source, title, url, published\_at, summary, raw). They standardize date formats, unify summary fields, and prepare both feeds for merging.

**Summary\_Check1**

**INPUT**

**OUTPUT**

**True Branch (270 items)**

**False Branch**

**Conditions**

/ {{ \$json.summary }} A is not empty

Add condition

Convert types where required

Options

No properties

Add option

**True Branch (270 items) Fields**

source	title	url	published_at	summary
Arxiv	LVLMs and Humans Ground Differently in Referential Communication	https://arxiv.org/abs/2601.19792	Fri, 30 Jan 2026 00:00:00 -0500	arXiv:2601.19792v2 Announce Type: replace-cross \nAbstract: For generative AI agents to partner effectively with human users, the ability to accurately predict human intent is critical. But this ability to collaborate remains limited by a critical deficit: an inability to model common ground. Here, we present a referential communication experiment with a factorial design involving director-matcher pairs (human-human, human-AI, AI-human, and AI-AI) that interact with multiple turns in repeated rounds to match pictures of objects not associated with any obvious lexicalized labels. We release the online pipeline for data collection, the tools and analyses for accuracy, efficiency, and lexical overlap, and a corpus of 356 dialogues (89 pairs over 4 rounds each) that unmasks LVLMs' limitations in interactively resolving referring expressions, a crucial skill that

**False Branch Fields**

source	title	url	published_at	summary

**d) Summary Validation Nodes:** These nodes remove records with empty or insufficient summaries, ensuring only meaningful content enters the final dataset. This enforces quality before limiting record counts.

**INPUT**

**Arxiv\_Limit60**

**Parameters**

Max Items: 100  
Keep: First Items

**OUTPUT**

100 items

source	title	url	published_at	summary	raw
Arxiv	LVLMs and Humans Ground Differently in Referential Communication	https://arxiv.org/abs/2601.19792	Fri, 30 Jan 00:00:00 -	arXiv:2601.19792v2 Announce Type: replace-cross \nAbstract: For generative AI agents to partner effectively with human users, the ability to accurately predict human intent is critical. But this ability to collaborate remains limited by a critical deficit: an inability to model common ground. Here, we present a referential communication experiment with a factorial design involving director-matcher pairs (human-human, human-AI, AI-human, and AI-AI) that interact with multiple turns in repeated rounds to match pictures of objects not associated with any obvious lexicalized labels. We release the online pipeline for data collection, the tools and analyses for accuracy, efficiency, and lexical overlap, and a corpus of 356 dialogues (89 pairs over 4 rounds each) that unmasks LVLMs' limitations in interactively resolving referring expressions, a crucial skill that	creator : Peter Zhenxiang Wang Dimitris Samaras Brennan, Owen R... rights : http://creativecommons.org/licenses/by/4.0/ title : LVLMs and Humans Ground Differently in Referential Communication link : https://arxiv.org/abs/2601.19792 pubDate : Fri, 30 Jan 00:00:00 -0500 content : arXiv:2601.19792v2 Announce Type: replace-cross \nAbstract: For generative AI agents to partner effectively with human users, the ability to accurately predict human intent is critical. But this ability to collaborate remains limited by a critical deficit: an inability to model common ground. Here, we present a referential communication experiment with a factorial design involving director-matcher pairs (human-human, human-AI, AI-human, and AI-AI) that interact with multiple turns in repeated rounds to match pictures of objects not associated with any obvious lexicalized labels. We release the online pipeline for data collection, the tools and analyses for accuracy, efficiency, and lexical overlap, and a corpus of 356 dialogues (89 pairs over 4 rounds each) that unmasks LVLMs' limitations in interactively resolving referring expressions, a crucial skill that

**e) Limit Nodes:** These nodes cap each RSS source to 60 clean records, enforcing quality-over-quantity and maintaining a balanced dataset size.

**INPUT**

**Parameter: "language"**

**Merge**

**Add Export Batch ID**

**Parameters**

Mode: Run Once for All Items

Language: JavaScript

```

1 const ts = new Date().toISOString().replace(/:/g, '-');
2 return input.all().map(i => {
3   json: { ...i.json, export_batch_id: ts }
4 });
5

```

**OUTPUT**

110 items

source	title	url	published_at	summary	raw
Arxiv	LVLMs and Humans Ground Differently in Referential Communication	https://arxiv.org/abs/2601.19792	Fri, 30 Jan 2026 00:00:00 -0500	arXiv:2601.19792v2 Announce Type: replace-cross \nAbstract: For generative AI agents to partner effectively with human users, the ability to accurately predict human intent is critical. But this ability to collaborate remains limited by a critical deficit: an inability to model common ground. Here, we present a referential communication experiment with a factorial design involving director-matcher pairs (human-human, human-AI, AI-human, and AI-AI) that interact with multiple turns in repeated rounds to match pictures of objects not associated with any obvious lexicalized labels. We release the online pipeline for data collection, the tools and analyses for accuracy, efficiency, and lexical overlap, and a corpus of 356 dialogues (89 pairs over 4 rounds each) that unmasks LVLMs' limitations in interactively resolving referring expressions, a crucial skill that	creator : Peter Zhenxiang Wang Dimitris Samaras Brennan, Owen R... rights : http://creativecommons.org/licenses/by/4.0/ title : LVLMs and Humans Ground Differently in Referential Communication link : https://arxiv.org/abs/2601.19792 pubDate : Fri, 30 Jan 00:00:00 -0500 content : arXiv:2601.19792v2 Announce Type: replace-cross \nAbstract: For generative AI agents to partner effectively with human users, the ability to accurately predict human intent is critical. But this ability to collaborate remains limited by a critical deficit: an inability to model common ground. Here, we present a referential communication experiment with a factorial design involving director-matcher pairs (human-human, human-AI, AI-human, and AI-AI) that interact with multiple turns in repeated rounds to match pictures of objects not associated with any obvious lexicalized labels. We release the online pipeline for data collection, the tools and analyses for accuracy, efficiency, and lexical overlap, and a corpus of 356 dialogues (89 pairs over 4 rounds each) that unmasks LVLMs' limitations in interactively resolving referring expressions, a crucial skill that

**f) Add Export Batch ID Node:** This node attaches a unique timestamp batch ID to each record, enabling traceability of exports and preventing confusion between multiple pipeline runs.

Parameter: "language"

INPUT

Schema Table JSON

Q Clean + Validate + Normalize Records Test step

Parameters Settings Docs

Mode: Run Once for All Items

Language: JavaScript

JavaScript:

```
1 const items = $input.all();
2
3 const seen = new Set();
4 const cutoff = new Date("2023-01-01");
5
6 const cleaned = [];
7
8 for (const i of items) {
9   const j = i.json();
10
11   const summary =
12     j.summary ||
13     j.raw2.content ||
14     j.raw2.description ||
15     "";
16
17   const url = j.url || j.link;
18   const publishedAt = j.published_at || j.isoDate || j.pubDate;
19
20   if (!url) continue;
21   if (!summary || summary.length <= 50) continue;
```

Type \$ for a list of special vars/methods. Debug by using `console.log()` statements and viewing their output in the browser console.

OUTPUT

110 items

source title url published\_at sum...

source	title	url	published_at	sum...
Arxiv	LVLMs and Humans Ground Differently in Referential Communication	<a href="https://arxiv.org/abs/2601.19792">https://arxiv.org/abs/2601.19792</a>	Fri, 30 Jan 2026 00:00:00 -0500	arXiv: \nAbstract: agents with h to acc intent to coll by a c to mo Here, comm with a involv pairs (

Arxiv LVLMs and Humans Ground Differently in Referential Communication

2026-01-30T05:00:00.000Z arXiv:  
\nAbstract:  
agents with h  
to acc  
intent  
to coll  
by a c  
to mo  
Here,  
comm  
with a  
involv  
pairs (

**g) Clean + Validate + Normalize Records Node:** This node performs final validation: removes duplicates by URL, filters outdated articles, standardizes publication dates to ISO-8601, consolidates raw provider payloads into a single column, and ensures summaries meet minimum length requirements.

- **Flow Execution** ([Screen Recording](#))
  - **Final CSV and JSON outputs**

row	title	url	published_at	summary	row	title	url
1	AKIN, and Human Sound Ground in Future War Communication	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00827">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00827</a>	2010-07-07T00:00:00Z	AKIN, and Human Sound Ground in Future War Communication Abstract: This article discusses the potential of sound as a weapon in future wars. The ability to detect potential human threat by a critical decision analysis is usually required in modern warfare. In addition, the potential of sound as a weapon in future wars is discussed. The potential of sound as a weapon in future wars is discussed.	10	"Peng, Wang, Li, and Peipei, Chengqiang Wang, Pengfei Kang, Qinmei Gao, Gegei Zhenhai, Susan Shuang, and Benji Zhou," <a href="https://doi.org/10.1002/1365-2788.00827">https://doi.org/10.1002/1365-2788.00827</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00827">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00827</a>
2	Pfeffer, Team-Level Approach for Federated Learning and Language Models	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00828">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00828</a>	2010-07-07T00:00:00Z	Pfeffer, Team-Level Approach for Federated Learning and Language Models Abstract: Federated learning has been widely adopted in various fields due to its ability to protect user privacy. However, the ability to detect potential human threat by a critical decision analysis is usually required in modern warfare. In addition, the potential of sound as a weapon in future wars is discussed.	11	"Wang, Wei, and Hanan Alshabani," <a href="https://doi.org/10.1002/1365-2788.00828">https://doi.org/10.1002/1365-2788.00828</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00828">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00828</a>
3	AKIN, Reactions, and Reflexes for UAV Performance Learning	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00829">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00829</a>	2010-07-07T00:00:00Z	AKIN, Reactions, and Reflexes for UAV Performance Learning Abstract: UAV performance learning is a challenging task due to the complex environment and the lack of prior knowledge. In this paper, we propose a novel approach for UAV performance learning based on the reactions and reflexes of the UAV. The proposed approach uses a reinforcement learning algorithm to learn the reactions and reflexes of the UAV, which are then used to improve the performance of the UAV.	12	"Dingling, Jiang, Zhou, Wei, and Yuan, Lin, and Jia, Lan, and Guo, Shuai, Peng, Yan," <a href="https://doi.org/10.1002/1365-2788.00829">https://doi.org/10.1002/1365-2788.00829</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00829">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00829</a>
4	Akpinar, Multi-Stage Model for First-Pass Metal Additive Manufacturing in Aeronautics CT	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00830">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00830</a>	2010-07-07T00:00:00Z	Akpinar, Multi-Stage Model for First-Pass Metal Additive Manufacturing in Aeronautics CT Abstract: In this study, a multi-stage model for first-pass metal additive manufacturing in aeronautics is proposed. The proposed model consists of three stages: pre-processing, processing, and post-processing. The pre-processing stage involves the generation of a digital model of the part, the selection of the appropriate manufacturing parameters, and the creation of a tool path. The processing stage involves the actual metal additive manufacturing process, which is carried out using a robotic arm. The post-processing stage involves the removal of any excess material and the final inspection of the part.	13	"Akpinar, Can, and Kavak, Omer," <a href="https://doi.org/10.1002/1365-2788.00830">https://doi.org/10.1002/1365-2788.00830</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00830">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00830</a>
5	IPO/FOA Aligning Small Clients with Their Free-Pass Preference for Knowledge Capital Question Answering	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00831">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00831</a>	2010-07-07T00:00:00Z	IPO/FOA Aligning Small Clients with Their Free-Pass Preference for Knowledge Capital Question Answering Abstract: In this paper, we propose a new approach for aligning small clients with their free-pass preference for knowledge capital question answering. The proposed approach uses a reinforcement learning algorithm to learn the free-pass preference of each client, which is then used to improve the performance of the knowledge capital question answering system.	14	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00831">https://doi.org/10.1002/1365-2788.00831</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00831">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00831</a>
6	A Systematic Inter-Module Relation for CodeQuality in Big Data Pipeline	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00832">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00832</a>	2010-07-07T00:00:00Z	A Systematic Inter-Module Relation for CodeQuality in Big Data Pipeline Abstract: In this paper, we propose a systematic inter-module relation for codequality in big data pipeline. The proposed approach uses a reinforcement learning algorithm to learn the inter-module relations between different modules, which are then used to improve the performance of the big data pipeline.	15	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00832">https://doi.org/10.1002/1365-2788.00832</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00832">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00832</a>
7	CIP-CDI Unsupervised Measures for Assessing Economic Condition	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00833">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00833</a>	2010-07-07T00:00:00Z	CIP-CDI Unsupervised Measures for Assessing Economic Condition Abstract: In this paper, we propose a new approach for assessing economic condition using CIP-CDI unsupervised measures. The proposed approach uses a reinforcement learning algorithm to learn the economic condition of a country, which is then used to improve the performance of the economic condition assessment system.	16	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00833">https://doi.org/10.1002/1365-2788.00833</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00833">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00833</a>
8	AYERET, Tali, Team-Tactical Review for New Fleet-Time Series Forecasting	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00834">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00834</a>	2010-07-07T00:00:00Z	AYERET, Tali, Team-Tactical Review for New Fleet-Time Series Forecasting Abstract: In this paper, we propose a new approach for team-tactical review for new fleet-time series forecasting. The proposed approach uses a reinforcement learning algorithm to learn the team-tactical review process, which is then used to improve the performance of the team-tactical review system.	17	"AYERET, Tali, and Tuncay, Tuncay," <a href="https://doi.org/10.1002/1365-2788.00834">https://doi.org/10.1002/1365-2788.00834</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00834">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00834</a>
9	LMI-Based H-infinity Filtering Problem: Realizing Program Termination Prediction	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00835">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00835</a>	2010-07-07T00:00:00Z	LMI-Based H-infinity Filtering Problem: Realizing Program Termination Prediction Abstract: In this paper, we propose a new approach for LMI-based H-infinity filtering problem: realizing program termination prediction. The proposed approach uses a reinforcement learning algorithm to learn the LMI-based H-infinity filtering problem, which is then used to improve the performance of the LMI-based H-infinity filtering problem.	18	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00835">https://doi.org/10.1002/1365-2788.00835</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00835">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00835</a>
10	The Generic Geometric Method for Linear Least Squares	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00836">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00836</a>	2010-07-07T00:00:00Z	The Generic Geometric Method for Linear Least Squares Abstract: In this paper, we propose a new approach for the generic geometric method for linear least squares. The proposed approach uses a reinforcement learning algorithm to learn the generic geometric method for linear least squares, which is then used to improve the performance of the generic geometric method for linear least squares.	19	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00836">https://doi.org/10.1002/1365-2788.00836</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00836">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00836</a>
11	From Speculative Generation, Utilitarian Learning to Fairness Learning on Unethical Medical Images	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00837">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00837</a>	2010-07-07T00:00:00Z	From Speculative Generation, Utilitarian Learning to Fairness Learning on Unethical Medical Images Abstract: In this paper, we propose a new approach for from speculative generation, utilitarian learning to fairness learning on unethical medical images. The proposed approach uses a reinforcement learning algorithm to learn the from speculative generation, utilitarian learning to fairness learning on unethical medical images, which is then used to improve the performance of the from speculative generation, utilitarian learning to fairness learning on unethical medical images.	20	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00837">https://doi.org/10.1002/1365-2788.00837</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00837">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00837</a>
12	Evacuation Time Adaptation Strategy for Efficient Evacuation	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00838">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00838</a>	2010-07-07T00:00:00Z	Evacuation Time Adaptation Strategy for Efficient Evacuation Abstract: In this paper, we propose a new approach for evacuation time adaptation strategy for efficient evacuation. The proposed approach uses a reinforcement learning algorithm to learn the evacuation time adaptation strategy for efficient evacuation, which is then used to improve the performance of the evacuation time adaptation strategy for efficient evacuation.	21	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00838">https://doi.org/10.1002/1365-2788.00838</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00838">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00838</a>
13	Utilization of Sensitivity FOMs for Cooperative Collision Prediction	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00839">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00839</a>	2010-07-07T00:00:00Z	Utilization of Sensitivity FOMs for Cooperative Collision Prediction Abstract: In this paper, we propose a new approach for utilization of sensitivity foms for cooperative collision prediction. The proposed approach uses a reinforcement learning algorithm to learn the utilization of sensitivity foms for cooperative collision prediction, which is then used to improve the performance of the utilization of sensitivity foms for cooperative collision prediction.	22	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00839">https://doi.org/10.1002/1365-2788.00839</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00839">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00839</a>
14	Scalability Loss-Free Path Planning for UAVs	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00840">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00840</a>	2010-07-07T00:00:00Z	Scalability Loss-Free Path Planning for UAVs Abstract: In this paper, we propose a new approach for scalability loss-free path planning for uav. The proposed approach uses a reinforcement learning algorithm to learn the scalability loss-free path planning for uav, which is then used to improve the performance of the scalability loss-free path planning for uav.	23	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00840">https://doi.org/10.1002/1365-2788.00840</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00840">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00840</a>
15	Qualitative Evaluation of the Effectiveness of the New Fuzzy Logic Language Model	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00841">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00841</a>	2010-07-07T00:00:00Z	Qualitative Evaluation of the Effectiveness of the New Fuzzy Logic Language Model Abstract: In this paper, we propose a new approach for qualitative evaluation of the effectiveness of the new fuzzy logic language model. The proposed approach uses a reinforcement learning algorithm to learn the qualitative evaluation of the effectiveness of the new fuzzy logic language model, which is then used to improve the performance of the qualitative evaluation of the effectiveness of the new fuzzy logic language model.	24	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00841">https://doi.org/10.1002/1365-2788.00841</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00841">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00841</a>
16	Pfeffer, Robust Decision-Making in Evaluating Risk-Free Fuzzy Logic Language Model	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00842">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00842</a>	2010-07-07T00:00:00Z	Pfeffer, Robust Decision-Making in Evaluating Risk-Free Fuzzy Logic Language Model Abstract: In this paper, we propose a new approach for robust decision-making in evaluating risk-free fuzzy logic language model. The proposed approach uses a reinforcement learning algorithm to learn the robust decision-making in evaluating risk-free fuzzy logic language model, which is then used to improve the performance of the robust decision-making in evaluating risk-free fuzzy logic language model.	25	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00842">https://doi.org/10.1002/1365-2788.00842</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00842">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00842</a>
17	Karakaya, and Pfeffer, Robust Decision-Making in Evaluating Risk-Free Fuzzy Logic Language Model	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00843">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00843</a>	2010-07-07T00:00:00Z	Karakaya, and Pfeffer, Robust Decision-Making in Evaluating Risk-Free Fuzzy Logic Language Model Abstract: In this paper, we propose a new approach for robust decision-making in evaluating risk-free fuzzy logic language model. The proposed approach uses a reinforcement learning algorithm to learn the robust decision-making in evaluating risk-free fuzzy logic language model, which is then used to improve the performance of the robust decision-making in evaluating risk-free fuzzy logic language model.	26	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00843">https://doi.org/10.1002/1365-2788.00843</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00843">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00843</a>
18	AKIN, Reactions, and Reflexes for UAV Performance Learning	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00844">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00844</a>	2010-07-07T00:00:00Z	AKIN, Reactions, and Reflexes for UAV Performance Learning Abstract: UAV performance learning is a challenging task due to the complex environment and the lack of prior knowledge. In this paper, we propose a novel approach for UAV performance learning based on the reactions and reflexes of the UAV. The proposed approach uses a reinforcement learning algorithm to learn the reactions and reflexes of the UAV, which are then used to improve the performance of the UAV.	27	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00844">https://doi.org/10.1002/1365-2788.00844</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00844">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00844</a>
19	AKIN, Reactions, and Reflexes for UAV Performance Learning	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00845">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00845</a>	2010-07-07T00:00:00Z	AKIN, Reactions, and Reflexes for UAV Performance Learning Abstract: UAV performance learning is a challenging task due to the complex environment and the lack of prior knowledge. In this paper, we propose a novel approach for UAV performance learning based on the reactions and reflexes of the UAV. The proposed approach uses a reinforcement learning algorithm to learn the reactions and reflexes of the UAV, which are then used to improve the performance of the UAV.	28	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00845">https://doi.org/10.1002/1365-2788.00845</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00845">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00845</a>
20	AKIN, Reactions, and Reflexes for UAV Performance Learning	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00846">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00846</a>	2010-07-07T00:00:00Z	AKIN, Reactions, and Reflexes for UAV Performance Learning Abstract: UAV performance learning is a challenging task due to the complex environment and the lack of prior knowledge. In this paper, we propose a novel approach for UAV performance learning based on the reactions and reflexes of the UAV. The proposed approach uses a reinforcement learning algorithm to learn the reactions and reflexes of the UAV, which are then used to improve the performance of the UAV.	29	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00846">https://doi.org/10.1002/1365-2788.00846</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00846">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00846</a>
21	AKIN, Reactions, and Reflexes for UAV Performance Learning	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00847">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00847</a>	2010-07-07T00:00:00Z	AKIN, Reactions, and Reflexes for UAV Performance Learning Abstract: UAV performance learning is a challenging task due to the complex environment and the lack of prior knowledge. In this paper, we propose a novel approach for UAV performance learning based on the reactions and reflexes of the UAV. The proposed approach uses a reinforcement learning algorithm to learn the reactions and reflexes of the UAV, which are then used to improve the performance of the UAV.	30	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00847">https://doi.org/10.1002/1365-2788.00847</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00847">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00847</a>
22	AKIN, Reactions, and Reflexes for UAV Performance Learning	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00848">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00848</a>	2010-07-07T00:00:00Z	AKIN, Reactions, and Reflexes for UAV Performance Learning Abstract: UAV performance learning is a challenging task due to the complex environment and the lack of prior knowledge. In this paper, we propose a novel approach for UAV performance learning based on the reactions and reflexes of the UAV. The proposed approach uses a reinforcement learning algorithm to learn the reactions and reflexes of the UAV, which are then used to improve the performance of the UAV.	31	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00848">https://doi.org/10.1002/1365-2788.00848</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00848">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00848</a>
23	AKIN, Reactions, and Reflexes for UAV Performance Learning	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00849">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00849</a>	2010-07-07T00:00:00Z	AKIN, Reactions, and Reflexes for UAV Performance Learning Abstract: UAV performance learning is a challenging task due to the complex environment and the lack of prior knowledge. In this paper, we propose a novel approach for UAV performance learning based on the reactions and reflexes of the UAV. The proposed approach uses a reinforcement learning algorithm to learn the reactions and reflexes of the UAV, which are then used to improve the performance of the UAV.	32	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00849">https://doi.org/10.1002/1365-2788.00849</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00849">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00849</a>
24	AKIN, Reactions, and Reflexes for UAV Performance Learning	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00850">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00850</a>	2010-07-07T00:00:00Z	AKIN, Reactions, and Reflexes for UAV Performance Learning Abstract: UAV performance learning is a challenging task due to the complex environment and the lack of prior knowledge. In this paper, we propose a novel approach for UAV performance learning based on the reactions and reflexes of the UAV. The proposed approach uses a reinforcement learning algorithm to learn the reactions and reflexes of the UAV, which are then used to improve the performance of the UAV.	33	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00850">https://doi.org/10.1002/1365-2788.00850</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00850">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00850</a>
25	AKIN, Reactions, and Reflexes for UAV Performance Learning	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00851">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00851</a>	2010-07-07T00:00:00Z	AKIN, Reactions, and Reflexes for UAV Performance Learning Abstract: UAV performance learning is a challenging task due to the complex environment and the lack of prior knowledge. In this paper, we propose a novel approach for UAV performance learning based on the reactions and reflexes of the UAV. The proposed approach uses a reinforcement learning algorithm to learn the reactions and reflexes of the UAV, which are then used to improve the performance of the UAV.	34	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00851">https://doi.org/10.1002/1365-2788.00851</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00851">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00851</a>
26	AKIN, Reactions, and Reflexes for UAV Performance Learning	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00852">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00852</a>	2010-07-07T00:00:00Z	AKIN, Reactions, and Reflexes for UAV Performance Learning Abstract: UAV performance learning is a challenging task due to the complex environment and the lack of prior knowledge. In this paper, we propose a novel approach for UAV performance learning based on the reactions and reflexes of the UAV. The proposed approach uses a reinforcement learning algorithm to learn the reactions and reflexes of the UAV, which are then used to improve the performance of the UAV.	35	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00852">https://doi.org/10.1002/1365-2788.00852</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00852">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00852</a>
27	AKIN, Reactions, and Reflexes for UAV Performance Learning	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00853">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00853</a>	2010-07-07T00:00:00Z	AKIN, Reactions, and Reflexes for UAV Performance Learning Abstract: UAV performance learning is a challenging task due to the complex environment and the lack of prior knowledge. In this paper, we propose a novel approach for UAV performance learning based on the reactions and reflexes of the UAV. The proposed approach uses a reinforcement learning algorithm to learn the reactions and reflexes of the UAV, which are then used to improve the performance of the UAV.	36	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00853">https://doi.org/10.1002/1365-2788.00853</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00853">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00853</a>
28	AKIN, Reactions, and Reflexes for UAV Performance Learning	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00854">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00854</a>	2010-07-07T00:00:00Z	AKIN, Reactions, and Reflexes for UAV Performance Learning Abstract: UAV performance learning is a challenging task due to the complex environment and the lack of prior knowledge. In this paper, we propose a novel approach for UAV performance learning based on the reactions and reflexes of the UAV. The proposed approach uses a reinforcement learning algorithm to learn the reactions and reflexes of the UAV, which are then used to improve the performance of the UAV.	37	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00854">https://doi.org/10.1002/1365-2788.00854</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00854">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00854</a>
29	AKIN, Reactions, and Reflexes for UAV Performance Learning	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00855">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00855</a>	2010-07-07T00:00:00Z	AKIN, Reactions, and Reflexes for UAV Performance Learning Abstract: UAV performance learning is a challenging task due to the complex environment and the lack of prior knowledge. In this paper, we propose a novel approach for UAV performance learning based on the reactions and reflexes of the UAV. The proposed approach uses a reinforcement learning algorithm to learn the reactions and reflexes of the UAV, which are then used to improve the performance of the UAV.	38	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00855">https://doi.org/10.1002/1365-2788.00855</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00855">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00855</a>
30	AKIN, Reactions, and Reflexes for UAV Performance Learning	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00856">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00856</a>	2010-07-07T00:00:00Z	AKIN, Reactions, and Reflexes for UAV Performance Learning Abstract: UAV performance learning is a challenging task due to the complex environment and the lack of prior knowledge. In this paper, we propose a novel approach for UAV performance learning based on the reactions and reflexes of the UAV. The proposed approach uses a reinforcement learning algorithm to learn the reactions and reflexes of the UAV, which are then used to improve the performance of the UAV.	39	"Yilmaz, Omer, and Yilmaz, Nihat," <a href="https://doi.org/10.1002/1365-2788.00856">https://doi.org/10.1002/1365-2788.00856</a>	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00856">https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00856</a>
31	AKIN, Reactions, and Reflexes for UAV Performance Learning	<a href="https://onlinelibrary.wiley.com/doi/10.1002/1365-2788.00857">https://onlinelibrary.wiley.com/doi/10.1002/1365-</a>					

[Arxiv & Smol RSS Feed CSV](#)

```

1  [
2    {
3      "collection_metadata": {
4        "dataset_name": "AlpacaEval Ground Truth Evaluations",
5        "source": "HuggingFace - tatsu-lab/alpaca_eval",
6        "source_url": "https://huggingface.co/datasets/tatsu-lab/alpaca_eval",
7        "purpose": "Ground truth instruction-response pairs for model behavior evaluation",
8        "use_case": "Agent 2 (Research Layer) uses these examples to compare model outputs against high-quality baselines",
9        "total_records": 80,
10       "collected_at": "2026-01-31T02:19:02.045Z",
11       "record_type": "ground_truth_evaluation",
12       "statistics": {
13         "total_records": 80,
14         "category_distribution": {
15           "explanation": 23,
16           "general": 34,
17           "instruction": 4,
18           "enumeration": 6,
19           "creative": 13
20         },
21         "dataset_distribution": {
22           "helpful_base": 16,
23           "koala": 16,
24           "oasst": 16,
25           "selfinstruct": 16,
26           "vicuna": 16
27         },
28         "avg_response_length": 638,
29         "model_used": "text_davinci_003",
30         "avg_quality_scores": {
31           "quality": 0.85,
32           "instruction_following": 0.9,
33           "coherence": 0.88,
34           "helpfulness": 0.87
35         }
36       },
37       "quality_metrics": {
38         "complete_records": 80,
39         "incomplete_records": 0,
40         "quality_rate": "100%",
41         "validation_checks": {
42           "all_prompts_present": true,
43           "all_responses_present": true,
44           "all_scores_valid": true,
45           "duplicates_removed": true,
46           "dates_standardized": true
47         }
48       },
49       "schema": {
50         "format": "flattened",
51         "nesting": "none",
52         "total_fields": 14,
53         "fields": [

```

### [AlpacaEval Dataset JSON](#)

- Replication Notes: A peer can replicate this pipeline by importing the workflow JSON, installing n8n locally, and executing once. No credentials are required.

## Final Output

The final step exports two datasets:

- `content_feed.csv` containing normalized RSS research and news
  - `alpaca_eval.json` containing flattened ground truth evaluations
- These files are saved locally and reused in Assignment 4 without re-running APIs.

**Both datasets are validated and ready for Assignment 4 agent integration.**

---