

AMTAIR Prototype Demonstration (Public Colab Notebook)

AMTAIR Prototype: Automating Transformative AI Risk Modeling

Executive Summary

This notebook implements a prototype of the AMTAIR (Automating Transformative AI Risk Modeling) project, which addresses the critical coordination failure in AI governance by developing computational tools that automate the extraction of probabilistic world models from AI safety literature.

The prototype demonstrates the transformation pipeline from structured argument representations (ArgDown) to probabilistic Bayesian networks (BayesDown), enabling the visualization and analysis of causal relationships and probability distributions that underlie AI risk assessments and policy evaluations.

Purpose Within the Master's Thesis

This notebook serves as the technical implementation component of the Master's thesis "Automating Transformative AI Risk Modeling: A Computational Approach to Policy Impact Evaluation." It demonstrates the feasibility of automating the extraction and formalization of world models, focusing on the core extraction pipeline and visualization capabilities that form the foundation for more sophisticated analysis.

Relevance to AI Governance

The coordination crisis in AI governance stems from different stakeholders working with incompatible assumptions, terminologies, and priorities. By making implicit models explicit through automated extraction and formalization, this work helps bridge communication gaps between technical researchers, policy specialists, and other stakeholders, contributing to more effective coordination in addressing existential risks from advanced AI.

Notebook Structure and Workflow

This notebook implements a multi-stage pipeline for transforming argument structures into interactive Bayesian network visualizations:

1. **Environment Setup** (Sections 0.1-0.3): Establishes the technical environment with necessary libraries and data connections
2. **Argument Extraction** (Sections 1.0-1.8): Processes source documents into structured ArgDown representations
3. **Probability Integration** (Sections 2.0-2.8): Enhances ArgDown with probability information to create BayesDown
4. **Data Transformation** (Section 3.0): Converts BayesDown into structured DataFrame format
5. **Visualization and Analysis** (Section 4.0): Creates interactive Bayesian network visualizations
6. **Archiving and Export** (Sections 5.0-6.0): Provides utilities for saving and sharing results

Throughout this notebook, we use the classic rain-sprinkler-lawn example as a canonical test case, demonstrating how a simple causal scenario (rain and sprinkler use affecting wet grass) can be represented, processed, and visualized using our automated pipeline.

Project Context and Purpose

This notebook implements a prototype of the Automating Transformative AI Risk Modeling (AMTAIR) project, which addresses a critical coordination failure in AI governance by developing computational tools to automate the extraction of probabilistic world models from AI safety literature.

The coordination crisis in AI governance stems from different stakeholders (technical researchers, policy specialists, ethicists) operating with different terminologies, priorities, and implicit theories of change. This fragmentation systematically increases existential risk through safety gaps, resource misallocation, and capability-governance mismatches.

The AMTAIR project aims to bridge these divides by: 1. Making implicit models explicit through automated extraction and formalization 2. Enabling comparison across different world-views 3. Providing a common language for discussing probabilistic relationships 4. Supporting policy evaluation across diverse scenarios

Notebook Overview and Pipeline

This notebook demonstrates the core extraction pipeline from structured argument representations (ArgDown) to probabilistic Bayesian networks (BayesDown), using the classic rain-sprinkler-lawn example as a canonical test case.

The pipeline consists of five main stages: 1. **Environment Setup**: Libraries, GitHub repository access, and data loading 2. **Argument Extraction**: Processing source documents into structured ArgDown format 3. **Probability Integration**: Enhancing ArgDown with probabilistic information to create BayesDown 4. **Data Transformation**: Converting BayesDown into structured DataFrame format 5. **Visualization & Analysis**: Creating interactive Bayesian network visualizations

Connection to Master's Thesis

This notebook serves as the technical implementation component of the Master's thesis "Automating Transformative AI Risk Modeling: A Computational Approach to Policy Impact Evaluation" (see PY_Thesis_OutlineNDraft), demonstrating the feasibility of automating the process of extracting and formalizing world models from AI safety literature.

The thesis positions this work as a solution to the coordination crisis in AI governance, where the AMTAIR tools provide a crucial bridge between different stakeholder communities by creating formal representations that can be analyzed, compared, and used for policy evaluation.

For broader context on the project's motivation and placement within AI governance efforts, see PY_Post0.0 ("The Missing Piece: Why We Need a Grand Strategy for AI") and PY_AMTAIRDescription, which explain how this technical work contributes to the development of a comprehensive AI safety grand strategy.

Instructions — How to use this notebook:

1. **Import Libraries & Install Packages**: Run Section 0.1 to set up the necessary dependencies for data processing and visualization.
2. **Connect to GitHub Repository & Load Data files**: Run Section 0.2 to establish connections to the data repository and load example datasets. This step retrieves sample ArgDown files and extracted data for demonstration.
3. **Process Source Documents to ArgDown**: Sections 1.0-1.8 demonstrate the extraction of argument structures from source documents (such as PDFs) into ArgDown format, a markdown-like notation for structured arguments.

4. **Convert ArgDown to BayesDown:** Sections 2.0-2.3 handle the transformation of ArgDown files into BayesDown format, which incorporates probabilistic information into the argument structure.
5. **Extract Data into Structured Format:** Section 3.0 processes BayesDown format into structured database entries (CSV) that can be used for analysis.
6. **Create and Analyze Bayesian Networks:** Section 4.0 demonstrates how to build Bayesian networks from the extracted data and provides tools for analyzing risk pathways.
7. **Save and Export Results:** Sections 5.0-6.0 provide methods for archiving results and exporting visualizations.

AMTAIR Prototype Demonstration (Public Colab Notebook)

AMTAIR Prototype: Automating Transformative AI Risk Modeling

Executive Summary

Purpose Within the Master's Thesis

Relevance to AI Governance

Notebook Structure and Workflow

Project Context and Purpose

Notebook Overview and Pipeline

Connection to Master's Thesis

Instructions — How to use this notebook:

Key Concepts:

Example Workflow:

Troubleshooting:

Environment Setup and Data Access

0.1 Prepare Colab/Python Environment — Import Libraries & Packages

0.2 Connect to GitHub Repository

0.3 File Import

1.0 Sources (PDF's of Papers) to ArgDown (.md file)

Sources to ArgDown: Structured Argument Extraction

Process Overview

What is ArgDown?

1.1 Specify Source Document (e.g. PDF)

1.2 Generate ArgDown Extraction Prompt

1.3 Prepare LLM API Call

1.4 Make ArgDown Extraction LLM API Call

1.5 Save ArgDown Extraction Response

1.6 Review and Check ArgDown.md File

1.6.2 Check the Graph Structure with the ArgDown Sandbox Online

1.7 Extract ArgDown Graph Information as DataFrame

1.8 Store ArgDown Information as ‘ArgDown.csv’ file

2.0 Probability Extractions: ArgDown (.csv) to BayesDown (.md + plugin JSON syntax)

ArgDown to BayesDown: Adding Probability Information

Process Overview

What is BayesDown?

2.1 Probability Extraction Questions — ‘ArgDown.csv’ to ‘ArgDown_WithQuestions.csv’

2.2 ‘ArgDown_WithQuestions.csv’ to ‘BayesDownQuestions.md’

2.3 Generate BayesDown Probability Extraction Prompt

2.3.1 BayesDown Format Specification

Core Structure

Rain-Sprinkler-Lawn Example

2.4 Prepare 2nd API call

2.5 Make BayesDown Probability Extraction API Call

2.6 Save BayesDown with Probability Estimates (.csv)

2.7	Review & Verify BayesDown Probability Estimates
2.7.2	Check the Graph Structure with the ArgDown Sandbox Online
2.8	Extract BayesDown with Probability Estimates as Dataframe
3.0	Data Extraction: BayesDown (.md) to Database (.csv)
BayesDown to Structured Data: Network Construction	
Extraction Pipeline Overview	
	Theoretical Foundation
	Role in Thesis Research
3.1	ExtractBayesDown-Data_v1
3.1.2	Test BayesDown Extraction
3.1.2.2	Check the Graph Structure with the ArgDown Sandbox Online
3.3	Extraction
3.3	Data-Post-Processing
3.4	Download and save finished data frame as .csv file
Analysis & Inference: Bayesian Network Visualization	
Bayesian Network Visualization Approach	
	Visualization Philosophy
	Connection to AMTAIR Goals
	Implementation Structure
Phase 1: Dependencies/Functions	
Phase 2: Node Classification and Styling Module	
Phase 3: HTML Content Generation Module	
Phase 4: Main Visualization Function	
Quickly check HTML Outputs	
Conclusion: From Prototype to Production	

Summary of Achievements

Limitations and Future Work

Connection to AMTAIR Project

6.0 Save Outputs

Saving and Exporting Results

Convert .ipynb Notebook to Markdown

Key Concepts:

- **ArgDown:** A structured format for representing arguments, with hierarchical relationships between statements.
- **BayesDown:** An extension of ArgDown that incorporates probabilistic information, allowing for Bayesian network construction.
- **Extraction Pipeline:** The process of converting unstructured text to structured argument representations.
- **Bayesian Networks:** Probabilistic graphical models that represent variables and their conditional dependencies.

Example Workflow:

1. Load a sample ArgDown file from the repository
2. Extract the hierarchical structure and relationships
3. Add probabilistic information to create a BayesDown representation
4. Generate a Bayesian network visualization
5. Analyze conditional probabilities and risk pathways

Troubleshooting:

- If connectivity issues occur, ensure you have access to the GitHub repository
- For visualization errors, check that all required libraries are properly installed
- When processing custom files, ensure they follow the expected format conventions

0. Environment Setup and Data Access

This section establishes the technical foundation for the AMTAIR prototype by: 1. Installing and importing necessary libraries 2. Setting up access to the GitHub repository 3. Loading example data files

The environment setup is designed to be run once per session, with flags to prevent redundant installations and imports. This section forms the basis for the subsequent extraction and analysis steps in the pipeline.

The key goal is to create a reproducible environment where the Bayesian network extraction and visualization can be performed consistently, with appropriate error handling and resource management.

0.1 Prepare Colab/Python Environment — Import Libraries & Packages

```
# @title 0.1 --- Install & Import Libraries & Packages (One-Time Setup) ---

"""
BLOCK PURPOSE: Establishes the core technical environment for the AMTAIR prototype.
Sets up all required libraries for Bayesian network processing, visualization, and data manipulation.
Uses a flag-based approach to ensure setup only runs once per session, enhancing efficiency.

The setup follows a three-stage process:
1. Install required packages not available in Colab by default
2. Import all necessary libraries with error handling
3. Set a global flag to prevent redundant execution

DEPENDENCIES: Requires internet connection for package installation
OUTPUTS: Global variable _setup_imports_done and loaded Python libraries
"""

# Check if setup has already been completed in this session using environment flag
try:
    # If this variable exists, setup was already done successfully
    _setup_imports_done
    print(" Libraries already installed and imported in this session. Skipping setup.")
except NameError:
```



```

print(" Performing one-time library installation and imports...")

# --- STAGE 1: Install required packages ---
# Install visualization and network analysis libraries
!pip install -q pyvis # Network visualization library
!apt-get install pandoc -y # Document conversion utility

# Install Google API and data processing packages
!pip install -q --upgrade gspread pandas google-auth google-colab # Data manipulation and processing

# Install Bayesian network and probabilistic modeling tools
!pip install -q pgmpy # Probabilistic graphical models library

# Install notebook conversion tools
!pip install -q nbconvert # Often pre-installed, but ensures availability

print(" --> Installations complete.")

# --- STAGE 2: Import libraries with error handling ---
try:
    # Network and HTTP libraries
    import requests # For making HTTP requests to APIs and GitHub
    import io # For handling in-memory file-like objects

    # Data processing libraries
    import pandas as pd # For structured data manipulation
    import numpy as np # For numerical operations
    import json # For JSON parsing and serialization
    import re # For regular expression pattern matching

    # Visualization libraries
    import matplotlib.pyplot as plt # For creating plots and charts
    from IPython.display import HTML, display, Markdown # For rich output in notebook

    # --- Specialized libraries requiring installation ---
    # Network analysis library
    import networkx as nx # For graph representation and analysis

    # Probabilistic modeling libraries
    from pgmpy.models import BayesianNetwork # For Bayesian network structure
    from pgmpy.factors.discrete import TabularCPD # For conditional probability tables
    from pgmpy.inference import VariableElimination # For probabilistic inference

```

```

# Interactive network visualization
from pyvis.network import Network # For interactive network visualization

# Output version information for key libraries
print(f"    pandas version: {pd.__version__}")
print(f"    networkx version: {nx.__version__}")
# Add others if specific versions are critical

print("    --> Imports complete.")

# --- STAGE 3: Set flag to indicate successful setup ---
_setup_imports_done = True
print("    One-time setup finished successfully.")

except ImportError as e:
    # Handle specific import failures
    print(f"    ERROR during import: {e}")
    print("    --> Setup did not complete successfully. Please check installations.")
except Exception as e:
    # Handle unexpected errors
    print(f"    UNEXPECTED ERROR during setup: {e}")
    print("    --> Setup did not complete successfully.")

# Environment is now ready for AMTAIR processing

```

:
 Running cells with 'Python 3.13.3' requires the ipykernel package.
<command:jupyter.createPythonEnvAndSelectController>>Create a Python Environment
 Or install 'ipykernel' using the command: '/opt/homebrew/bin/python3 -m pip install ipykernel'

0.2 Connect to GitHub Repository

The Public GitHub Repo Url in use:

https://raw.githubusercontent.com/SingularitySmith/AMTAIR_Prototype/main/

Note: When encountering errors, accessing the data, try using “RAW” Urls.

```

# @title 0.2 --- Connect to GitHub Repository --- Load Files

"""
BLOCK PURPOSE: Establishes connection to the AMTAIR GitHub repository and provides

```

functions to load example data files for processing.

This block creates a reusable function for accessing files from the project's GitHub repository, enabling access to example files like the rain-sprinkler-lawn Bayesian network that serves as our canonical test case.

DEPENDENCIES: requests library, io library

OUTPUTS: load_file_from_repo function and test file loads

"""

```
from requests.exceptions import HTTPError
```

```
# Specify the base repository URL for the AMTAIR project
```

```
repo_url = "https://raw.githubusercontent.com/SingularitySmith/AMTAIR_Prototype/main/data/ex"
```

```
print(f"Connecting to repository: {repo_url}")
```

```
def load_file_from_repo(relative_path):
```

```
    """
```

```
    Loads a file from the specified GitHub repository using a relative path.
```

```
    Args:
```

```
        relative_path (str): Path to the file relative to the repo_url
```

```
    Returns:
```

```
        For CSV/JSON: pandas DataFrame
```

```
        For MD: string containing file contents
```

```
    Raises:
```

```
        HTTPError: If file not found or other HTTP error occurs
```

```
        ValueError: If unsupported file type is requested
```

```
    """
```

```
    file_url = repo_url + relative_path
```

```
    print(f"Attempting to load: {file_url}")
```

```
    # Fetch the file content from GitHub
```

```
    response = requests.get(file_url)
```

```
    # Check for bad status codes with enhanced error messages
```

```
    if response.status_code == 404:
```

```
        raise HTTPError(f"File not found at URL: {file_url}. Check the file path/name and en"
```

```
    else:
```

```
        response.raise_for_status() # Raise for other error codes
```

```

# Convert response to file-like object
file_object = io.StringIO(response.text)

# Process different file types appropriately
if relative_path.endswith(".csv"):
    return pd.read_csv(file_object) # Return DataFrame for CSV
elif relative_path.endswith(".json"):
    return pd.read_json(file_object) # Return DataFrame for JSON
elif relative_path.endswith(".md"):
    return file_object.read() # Return raw content for MD files
else:
    raise ValueError(f"Unsupported file type: {relative_path.split('.')[-1]}. Add support")

# Load example files to test connection
try:
    # Load the extracted data CSV file
    # df = load_file_from_repo("extracted_data.csv")

    # Load the ArgDown test text
    md_content = load_file_from_repo("ArgDown.md")

    print(" Successfully connected to repository and loaded test files.")
except Exception as e:
    print(f" Error loading files: {str(e)}")
    print("Please check your internet connection and the repository URL.")

# Display preview of loaded content (commented out to avoid cluttering output)
print(md_content)

```

0.3 File Import

```

# @title
md_content

```

1.0 Sources (PDF's of Papers) to ArgDown (.md file)

1. Sources to ArgDown: Structured Argument Extraction

Process Overview

This section implements the first major stage of the AMTAIR pipeline: transforming source documents (such as research papers, blog posts, or expert analyses) into structured argument representations using the ArgDown format.

ArgDown is a markdown-like notation for representing arguments in a hierarchical structure. In the context of AMTAIR, it serves as the first step toward creating formal Bayesian networks by:

1. Identifying key variables/statements in the text
2. Capturing their hierarchical relationships
3. Preserving their descriptive content
4. Defining their possible states (instantiations)

The extraction process uses Large Language Models (LLMs) to identify the structure and relationships in the text, though in this notebook we focus on processing pre-formatted examples rather than performing the full extraction from raw text.

What is ArgDown?

ArgDown uses a simple syntax where:

- Statements are represented as `[Statement]: Description`
- Relationships are indicated with `+` symbols and indentation
- Metadata is added in JSON format, including possible states of each variable

For example:

```
[MainClaim]: Description of the main claim. {"instantiations": ["claim_TRUE", "claim_FALSE"]}
+ [SupportingEvidence]: Description of evidence. {"instantiations": ["evidence_TRUE", "evidence_FALSE"]}
```

This structure will later be enhanced with probability information to create BayesDown, which can be transformed into a Bayesian network for analysis and visualization.

1.1 Specify Source Document (e.g. PDF)

Review the source document, ensure it is suitable for API call and upload to / store it in the correct location.

```
# @title 1.1.a) --- MTAIR Online Model (Analytica) ---
```

```
from IPython.display import IFrame
```

```
IFrame(src="https://acp.analytica.com/view0?invite=4560&code=3000289064591444815", width="1000px", height="500px")
```

```
<IPython.lib.display.IFrame at 0x7d3db6f27310>
```

1.2 Generate ArgDown Extraction Prompt

Generate Extraction Prompt

```
# @title 1.2.0 --- Prompt Template Function Definitions ---
```

```
"""
```

```
BLOCK PURPOSE: Defines a flexible template system for LLM prompts used in the extraction pipeline
```

```
This block implements two key classes:
```

1. PromptTemplate: A simple template class supporting variable substitution for dynamic prompts
2. PromptLibrary: A collection of pre-defined prompt templates for different extraction tasks

```
These templates are used in the ArgDown extraction and BayesDown probability extraction stages of the pipeline, providing consistent and well-structured prompts to the LLMs.
```

```
DEPENDENCIES: string.Template for variable substitution
```

```
OUTPUTS: PromptTemplate and PromptLibrary classes
```

```
"""
```

```
from string import Template
```

```
from typing import Dict, Optional, Union, List
```

```
class PromptTemplate:
```

```
    """Template system for LLM prompts with variable substitution"""
```

```
    def __init__(self, template: str):
```

```
        """Initialize with template string using $variable format"""
```

```
        self.template = Template(template)
```

```
    def format(self, **kwargs) -> str:
```

```
        """Substitute variables in the template"""
```

```
        return self.template.safe_substitute(**kwargs)
```

```

    @classmethod
    def from_file(cls, filepath: str) -> 'PromptTemplate':
        """Load template from a file"""
        with open(filepath, 'r') as f:
            template = f.read()
        return cls(template)

class PromptLibrary:
    """Collection of prompt templates for different extraction tasks"""

    # ArgDown extraction prompt - transforms source text into structured argument map
    ARGDOWN_EXTRACTION = PromptTemplate("""
You are participating in the AMTAIR (Automating Transformative AI Risk Modeling) project and
Your specific task is to extract the implicit causal model from the provided document in stru

## Epistemic Foundation & Purpose

This extraction represents one possible interpretation of the implicit causal model in the d

Your role is to reveal the causal structure already present in the author's thinking, mainta

## ArgDown Format Specification

### Core Syntax

ArgDown represents causal relationships using a hierarchical structure:

1. Variables appear in square brackets with descriptive text:
    `[Variable_Name]: Description of the variable.`

2. Causal relationships use indentation (2 spaces per level) and '+' symbols:

    [Effect]: Description of effect. + [Cause]: Description of cause. + [Deeper_Cause]: Descript

3. Causality flows from bottom (more indented) to top (less indented):
    - More indented variables (causes) influence less indented variables (effects)
    - The top-level variable is the ultimate effect or outcome
    - Deeper indentation levels represent root causes or earlier factors

4. Each variable must include JSON metadata with possible states (instantiations):
    `[Variable]: Description. {"instantiations": ["variable_STATE1", "variable_STATE2"]}`

```

JSON Metadata Format

The JSON metadata must follow this exact structure:

```
```json
{"instantiations": ["variable_STATE1", "variable_STATE2"]}
```

Requirements:

- \* Double quotes (not single) around field names and string values
- \* Square brackets enclosing the instantiations array
- \* Comma separation between array elements
- \* No trailing comma after the last element
- \* Must be valid JSON syntax that can be parsed by standard JSON parsers

For binary variables (most common case):

```
{"instantiations": ["variable_TRUE", "variable_FALSE"]}
```

For multi-state variables (when clearly specified in the text):

```
{"instantiations": ["variable_HIGH", "variable_MEDIUM", "variable_LOW"]}
```

The metadata must appear on the same line as the variable definition, after the description.

### ## Complex Structural Patterns

#### ### Variables Influencing Multiple Effects

The same variable can appear multiple times in different places in the hierarchy if it influences multiple effects.

```
[Effect1]: First effect description. {"instantiations": ["effect1_TRUE", "effect1_FALSE"]}
+ [Cause_A]: Description of cause A. {"instantiations": ["cause_a_TRUE", "cause_a_FALSE"]}
```

```
[Effect2]: Second effect description. {"instantiations": ["effect2_TRUE", "effect2_FALSE"]}
+ [Cause_A]
+ [Cause_B]: Description of cause B. {"instantiations": ["cause_b_TRUE", "cause_b_FALSE"]}
```

#### ### Multiple Causes of the Same Effect

Multiple causes can influence the same effect by being listed at the same indentation level:

```
[Effect]: Description of effect. {"instantiations": ["effect_TRUE", "effect_FALSE"]}
+ [Cause1]: Description of first cause. {"instantiations": ["cause1_TRUE", "cause1_FALSE"]}
+ [Cause2]: Description of second cause. {"instantiations": ["cause2_TRUE", "cause2_FALSE"]}
+ [Deeper_Cause]: A cause that influences Cause2. {"instantiations": ["deeper_cause_TRUE", "deeper_cause_FALSE"]}
```

#### ### Causal Chains

Causal chains are represented through multiple levels of indentation:

```
[Ultimate_Effect]: The final outcome. {"instantiations": ["ultimate_effect_TRUE", "ultimate_effect_FALSE"]}
+ [Intermediate_Effect]: A mediating variable. {"instantiations": ["intermediate_effect_TRUE", "intermediate_effect_FALSE"]}
```



```

+ [Root_Cause]: The initial cause. {"instantiations": ["root_cause_TRUE", "root_cause_FALSE"]}
+ [2nd_Intermediate_Effect]: A mediating variable. {"instantiations": ["intermediate_effect_TRUE", "intermediate_effect_FALSE"]}

Common Cause of Multiple Variables
A common cause affecting multiple variables is represented by referencing the same variable :
[Effect1]: First effect description. {"instantiations": ["effect1_TRUE", "effect1_FALSE"]}
+ [Common_Cause]: Description of common cause. {"instantiations": ["common_cause_TRUE", "common_cause_FALSE"]}

[Effect2]: Second effect description. {"instantiations": ["effect2_TRUE", "effect2_FALSE"]}
+ [Common_Cause]

Detailed Extraction Workflow
Please follow this step-by-step process, documenting your reasoning in XML tags:
<analysis>
First, conduct a holistic analysis of the document:
1. Identify the main subject matter or domain
2. Note key concepts, variables, and factors discussed
3. Pay attention to language indicating causal relationships (causes, affects, influences, depends on)
4. Look for the ultimate outcomes or effects that are the focus of the document
5. Record your general understanding of the document's implicit causal structure
</analysis>
<variable_identification>
Next, identify and list the key variables in the causal model:
* Focus on factors that are discussed as having an influence or being influenced
* For each variable:
 * Create a descriptive name in [square_brackets]
 * Write a concise description based directly on the text
 * Determine possible states (usually binary TRUE/FALSE unless clearly specified)
* Distinguish between:
 * Outcome variables (effects the author is concerned with)
 * Intermediate variables (both causes and effects in chains)
 * Root cause variables (exogenous factors in the model)
* List all identified variables with their descriptions and possible states
</variable_identification>

<causal_structure>
Then, determine the causal relationships between variables:
* For each variable, identify what factors influence it
* Note the direction of causality (what causes what)
* Look for mediating variables in causal chains
* Identify common causes of multiple effects

```

```
* Capture feedback loops if present (though they must be represented as DAGs)
* Map out the hierarchical structure of the causal model
</causal_structure>
```

```
<format_conversion>
```

Now, convert your analysis into proper ArgDown format:

```
* Start with the ultimate outcome variables at the top level
* Place direct causes indented below with \+ symbols
* Continue with deeper causes at further indentation levels
* Add variable descriptions and instantiations metadata
* Ensure variables appearing in multiple places have consistent names
* Check that the entire structure forms a valid directed acyclic graph
</format_conversion>
```

```
<validation>
```

Finally, review your extraction for quality and format correctness:

1. Verify all variables have properly formatted metadata
2. Check that indentation properly represents causal direction
3. Confirm the extraction accurately reflects the document's implicit model
4. Ensure no cycles exist in the causal structure
5. Verify that variables referenced multiple times are consistent
6. Check that the extraction would be useful for subsequent analysis

```
</validation>
```

### ## Source Document Analysis Guidance

When analyzing the source document:

- \* Focus on revealing the author's own causal model, not imposing an external framework
- \* Maintain the author's terminology where possible
- \* Look for both explicit statements of causality and implicit assumptions
- \* Pay attention to the relative importance the author assigns to different factors
- \* Notice where the author expresses certainty versus uncertainty
- \* Consider the level of granularity appropriate to the document's own analysis

Remember that your goal is to make the implicit model explicit, not to evaluate or improve it. The value lies in accurately representing the author's perspective, even if you might personally disagree.

```
""")
```

```
BayesDown probability extraction prompt - enhances ArgDown with probability information
```

```

BAYESDOWN_EXTRACTION = PromptTemplate("""
You are an expert in probabilistic reasoning and Bayesian networks. Your task is to extend the
ArgDown structure with probabilities.

For each statement in the ArgDown structure, you need to:
1. Estimate prior probabilities for each possible state
2. Estimate conditional probabilities given parent states
3. Maintain the original structure and relationships

Here is the format to follow:
[Node]: Description. { "instantiations": ["node_TRUE", "node_FALSE"], "priors": { "p(node_TRUE)"
[Parent]: Parent description. {...}

Here are the specific probability questions to answer:
$questions

ArgDown structure to enhance:
$argdown

Provide the complete BayesDown representation with probabilities:
""")

@classmethod
def get_template(cls, template_name: str) -> PromptTemplate:
 """Get a prompt template by name"""
 if hasattr(cls, template_name):
 return getattr(cls, template_name)
 else:
 raise ValueError(f"Template not found: {template_name}")

```

### 1.3 Prepare LLM API Call

Combine Systemprompt + API Specifications + ArgDown Instructions + Prompt + Source PDF for API Call

```

@title 1.3.0 --- Provider-Agnostic LLM API Interface ---

"""
BLOCK PURPOSE: Provides a unified interface for interacting with different LLM providers.

This block implements a flexible, provider-agnostic system for making LLM API calls:

```

1. Base abstract class (LLMProvider) defining the common interface
2. Implementation classes for specific providers (OpenAI and Anthropic)
3. Factory class for creating appropriate provider instances

This abstraction allows the extraction pipeline to work with different LLM providers without changing the core code, supporting both current and future LLM backends.

DEPENDENCIES: requests for API calls, os for environment variables, abstract base classes  
 OUTPUTS: LLMProvider abstract class and concrete implementations for OpenAI and Anthropic  
 """

```
import os
import json
import time
import requests
from abc import ABC, abstractmethod
from typing import Dict, List, Optional, Union, Any
from dataclasses import dataclass

@dataclass
class LLMResponse:
 """Standard response object for LLM completions"""
 content: str # The generated text response
 model: str # The model used for generation
 usage: Dict[str, int] # Token usage statistics
 raw_response: Dict[str, Any] # Complete provider-specific response
 created_at: float = time.time() # Timestamp of response creation

class LLMProvider(ABC):
 """Abstract base class for LLM providers"""

 @abstractmethod
 def complete(self,
 prompt: str,
 system_prompt: Optional[str] = None,
 temperature: float = 0.7,
 max_tokens: int = 4000) -> LLMResponse:
 """Generate a completion from the LLM"""
 pass

 @abstractmethod
 def get_available_models(self) -> List[str]:
```

```

 """Return a list of available models from this provider"""
 pass

class OpenAIProvider(LLMProvider):
 """OpenAI API implementation"""

 def __init__(self, api_key: Optional[str] = None, organization: Optional[str] = None):
 """Initialize with API key from args or environment"""
 self.api_key = api_key or os.environ.get("OPENAI_API_KEY")
 if not self.api_key:
 raise ValueError("OpenAI API key is required. Provide as argument or set OPENAI_API_KEY")

 self.organization = organization or os.environ.get("OPENAI_ORGANIZATION")
 self.api_base = "https://api.openai.com/v1"

 def complete(self,
 prompt: str,
 system_prompt: Optional[str] = None,
 model: str = "gpt-4-turbo",
 temperature: float = 0.7,
 max_tokens: int = 4000) -> LLMResponse:
 """Generate a completion using OpenAI's API"""

 # Prepare request headers
 headers = {
 "Content-Type": "application/json",
 "Authorization": f"Bearer {self.api_key}"
 }

 if self.organization:
 headers["OpenAI-Organization"] = self.organization

 # Create message structure
 messages = []
 if system_prompt:
 messages.append({"role": "system", "content": system_prompt})

 messages.append({"role": "user", "content": prompt})

 # Prepare request data
 data = {
 "model": model,

```

```

 "messages": messages,
 "temperature": temperature,
 "max_tokens": max_tokens
 }

 # Make API call
 response = requests.post(
 f"{self.api_base}/chat/completions",
 headers=headers,
 json=data
)

 response.raise_for_status()
 result = response.json()

 # Transform into standardized response format
 return LLMResponse(
 content=result["choices"][0]["message"]["content"],
 model=result["model"],
 usage=result["usage"],
 raw_response=result
)

def get_available_models(self) -> List[str]:
 """Return a list of available OpenAI models"""
 headers = {
 "Authorization": f"Bearer {self.api_key}"
 }

 if self.organization:
 headers["OpenAI-Organization"] = self.organization

 response = requests.get(
 f"{self.api_base}/models",
 headers=headers
)

 response.raise_for_status()
 models = response.json()["data"]
 return [model["id"] for model in models]

class AnthropicProvider(LLMProvider):

```

```

"""Anthropic Claude API implementation"""

def __init__(self, api_key: Optional[str] = None):
 """Initialize with API key from args or environment"""
 self.api_key = api_key or os.environ.get("ANTHROPIC_API_KEY")
 if not self.api_key:
 raise ValueError("Anthropic API key is required. Provide as argument or set ANTHROPIC_API_KEY")

 self.api_base = "https://api.anthropic.com/v1"

def complete(self,
 prompt: str,
 system_prompt: Optional[str] = None,
 model: str = "claude-3-opus-20240229",
 temperature: float = 0.7,
 max_tokens: int = 4000) -> LLMResponse:
 """Generate a completion using Anthropic's API"""

 # Prepare request headers
 headers = {
 "Content-Type": "application/json",
 "X-API-Key": self.api_key,
 "anthropic-version": "2023-06-01"
 }

 # Prepare request data in Anthropic-specific format
 data = {
 "model": model,
 "messages": [{"role": "user", "content": prompt}],
 "temperature": temperature,
 "max_tokens": max_tokens
 }

 # Add system prompt if provided (Anthropic uses a different format)
 if system_prompt:
 data["system"] = system_prompt

 # Make API call
 response = requests.post(
 f"{self.api_base}/messages",
 headers=headers,
 json=data
)

```

```

)

 response.raise_for_status()
 result = response.json()

 # Transform into standardized response format
 return LLMResponse(
 content=result["content"][0]["text"],
 model=result["model"],
 usage={"prompt_tokens": result.get("usage", {}).get("input_tokens", 0),
 "completion_tokens": result.get("usage", {}).get("output_tokens", 0)},
 raw_response=result
)

def get_available_models(self) -> List[str]:
 """Return a list of available Anthropic models"""
 # Anthropic doesn't have a models endpoint, so we return a static list
 return [
 "claude-3-opus-20240229",
 "claude-3-sonnet-20240229",
 "claude-3-haiku-20240307"
]

class LLMFactory:
 """Factory for creating LLM providers"""

 @staticmethod
 def create_provider(provider_name: str, **kwargs) -> LLMProvider:
 """Create and return an LLM provider instance"""
 if provider_name.lower() == "openai":
 return OpenAIProvider(**kwargs)
 elif provider_name.lower() == "anthropic":
 return AnthropicProvider(**kwargs)
 else:
 raise ValueError(f"Unsupported provider: {provider_name}")

@title 1.3.0 --- API Call Function Definitions ---

"""
BLOCK PURPOSE: Provides core functions for extracting ArgDown representations from text using

This block implements the main extraction functionality:

```



1. `extract_argdown_from_text`: Sends text to LLM to extract structured ArgDown representation
2. `validate_argdown`: Verifies the extracted ArgDown for correctness and completeness
3. `process_source_document`: Handles source files (PDF, TXT, MD) and manages extraction
4. `save_argdown_extraction`: Saves extraction results with metadata for further processing

These functions form the first stage of the AMTAIR pipeline, transforming unstructured text into structured argument representations.

DEPENDENCIES: LLMFactory from previous cell, re for pattern matching

OUTPUTS: Functions for ArgDown extraction, validation, and storage

"""

```
def extract_argdown_from_text(text: str, provider_name: str = "openai", model: str = None) ->
```

"""

Extract ArgDown representation from text using LLM

Args:

text: The source text to extract arguments from

provider\_name: The LLM provider to use (openai or anthropic)

model: Specific model to use, or None for default

Returns:

Extracted ArgDown representation

"""

# Create LLM provider

provider = LLMFactory.create\_provider(provider\_name)

# Get extraction prompt

prompt\_template = PromptLibrary.get\_template("ARGDOWN\_EXTRACTION")

prompt = prompt\_template.format(text=text)

# Set model-specific parameters

if provider\_name.lower() == "openai":

model = model or "gpt-4-turbo"

temperature = 0.3 # Lower temperature for more deterministic extraction

max\_tokens = 4000

elif provider\_name.lower() == "anthropic":

model = model or "claude-3-opus-20240229"

temperature = 0.2

max\_tokens = 4000

# Call the LLM

```

system_prompt = "You are an expert in argument mapping and causal reasoning."
response = provider.complete(
 prompt=prompt,
 system_prompt=system_prompt,
 model=model,
 temperature=temperature,
 max_tokens=max_tokens
)

Extract the ArgDown content (remove any markdown code blocks if present)
argdown_content = response.content
if "```" in argdown_content:
 # Extract content between code blocks if present
 import re
 matches = re.findall(r"```(?:argdown)?\n([\s\S]*?)\n```", argdown_content)
 if matches:
 argdown_content = matches[0]

return argdown_content

def validate_argdown(argdown_text: str) -> Dict[str, Any]:
 """
 Validate ArgDown representation to ensure it's well-formed

 Args:
 argdown_text: ArgDown representation to validate

 Returns:
 Dictionary with validation results
 """
 # Initialize validation results
 results = {
 "is_valid": True,
 "errors": [],
 "warnings": [],
 "stats": {
 "node_count": 0,
 "relationship_count": 0,
 "max_depth": 0
 }
 }

```

```

Basic syntax checks
lines = argdown_text.split("\n")
node_pattern = r'\[(.*?)\]:'
instantiation_pattern = r'{"instantiations":'

Track nodes and relationships
nodes = set()
relationships = []
current_depth = 0
max_depth = 0

for i, line in enumerate(lines):
 # Skip empty lines
 if not line.strip():
 continue

 # Calculate indentation depth
 indent = 0
 if '+' in line:
 indent = line.find('+') // 2

 current_depth = indent
 max_depth = max(max_depth, current_depth)

 # Check for node definitions
 import re
 node_matches = re.findall(node_pattern, line)
 if node_matches:
 node = node_matches[0]
 nodes.add(node)
 results["stats"]["node_count"] += 1

 # Check for instantiations
 if instantiation_pattern not in line:
 results["warnings"].append(f"Line {i+1}: Node '{node}' is missing instantiation")

 # Check parent-child relationships
 if indent > 0 and '+' in line and node_matches:
 # This is a child node; find its parent
 parent_indent = indent - 1
 j = i - 1
 while j >= 0:

```

```

 if '+' in lines[j] and lines[j].find('+') // 2 == parent_indent:
 parent_matches = re.findall(node_pattern, lines[j])
 if parent_matches:
 parent = parent_matches[0]
 relationships.append((parent, node))
 results["stats"]["relationship_count"] += 1
 break
 j -= 1

 results["stats"]["max_depth"] = max_depth

 # If we didn't find any nodes, that's a problem
 if results["stats"]["node_count"] == 0:
 results["is_valid"] = False
 results["errors"].append("No valid nodes found in ArgDown representation")

 return results

def process_source_document(file_path: str, provider_name: str = "openai") -> Dict[str, Any]:
 """
 Process a source document to extract ArgDown representation

 Args:
 file_path: Path to the source document
 provider_name: The LLM provider to use

 Returns:
 Dictionary with extraction results
 """
 # Load the source document
 text = ""
 if file_path.endswith(".pdf"):
 # PDF handling requires additional libraries
 try:
 import PyPDF2
 with open(file_path, 'rb') as file:
 reader = PyPDF2.PdfReader(file)
 text = ""
 for page in reader.pages:
 text += page.extract_text() + "\n"
 except ImportError:
 raise ImportError("PyPDF2 is required for PDF processing. Install it with: pip install PyPDF2")

```

```

elif file_path.endswith(".txt"):
 with open(file_path, 'r') as file:
 text = file.read()
elif file_path.endswith(".md"):
 with open(file_path, 'r') as file:
 text = file.read()
else:
 raise ValueError(f"Unsupported file format: {file_path}")

Extract ArgDown
argdown_content = extract_argdown_from_text(text, provider_name)

Validate the extraction
validation_results = validate_argdown(argdown_content)

Prepare results
results = {
 "source_path": file_path,
 "extraction_timestamp": time.time(),
 "argdown_content": argdown_content,
 "validation": validation_results,
 "provider": provider_name
}

return results

def save_argdown_extraction(results: Dict[str, Any], output_path: str) -> None:
 """
 Save ArgDown extraction results

 Args:
 results: Extraction results dictionary
 output_path: Path to save the results
 """
 # Save the ArgDown content
 with open(output_path, 'w') as file:
 file.write(results["argdown_content"])

 # Save metadata alongside
 metadata_path = output_path.replace('.md', '_metadata.json')
 metadata = {
 "source_path": results["source_path"],

```

```

 "extraction_timestamp": results["extraction_timestamp"],
 "validation": results["validation"],
 "provider": results["provider"]
 }

```

```

with open(metadata_path, 'w') as file:
 json.dump(metadata, file, indent=2)

```

```

@title 1.3 --- Prepare LLM API Call ---

```

```

"""

```

BLOCK PURPOSE: Prepares parameters for LLM API calls used in ArgDown extraction.

This function handles the configuration for LLM API calls, including:

1. Source document path validation
2. LLM provider selection and validation
3. Model selection with appropriate defaults

The function returns a configuration dictionary that can be passed to the extraction function in the next step of the pipeline.

DEPENDENCIES: None (uses standard Python functionality)

OUTPUTS: Dictionary with extraction configuration parameters

```

"""

```

```

def prepare_extraction_call(source_path, provider_name="openai", model=None):
 """

```

Prepare the LLM API call for ArgDown extraction

Args:

source\_path (str): Path to the source document to extract from  
 provider\_name (str): LLM provider to use ('openai' or 'anthropic')  
 model (str, optional): Specific model to use. Defaults to None (uses provider's default)

Returns:

dict: Configuration parameters for extraction

Raises:

ValueError: If an unsupported provider is specified

```

"""

```

# Load the source document

```

print(f"Processing source document: {source_path}")

```

```

Determine provider and model
provider = provider_name.lower()
if provider not in ["openai", "anthropic"]:
 raise ValueError(f"Unsupported provider: {provider}. Use 'openai' or 'anthropic'.")

Set default model if none provided
if model is None:
 if provider == "openai":
 model = "gpt-4-turbo"
 elif provider == "anthropic":
 model = "claude-3-opus-20240229"

Print configuration
print(f"Using provider: {provider}")
print(f"Selected model: {model}")

return {
 "source_path": source_path,
 "provider": provider,
 "model": model
}

Usage example:
source_path = "example_document.pdf" # Replace with actual document path
extraction_config = prepare_extraction_call(source_path, provider_name="openai")

```

Processing source document: example\_document.pdf  
Using provider: openai  
Selected model: gpt-4-turbo

## 1.4 Make ArgDown Extraction LLM API Call

```

@title 1.4 --- Make ArgDown Extraction LLM API Call ---

"""
BLOCK PURPOSE: Executes the ArgDown extraction process using the LLM API.

This function performs the actual extraction of ArgDown representations from source documents.
1. Takes the configuration parameters prepared in the previous step

```

2. Processes the document using the LLM API
3. Validates the extraction results
4. Provides timing and statistics about the extraction

The extraction process transforms unstructured text into a structured argument representation following the ArgDown syntax defined in the AMTAIR project.

DEPENDENCIES: process\_source\_document function from previous cells

OUTPUTS: Dictionary with extraction results including ArgDown content and validation info  
 """

```
def execute_extraction(extraction_config):
 """
 Execute the ArgDown extraction using the LLM API

 Args:
 extraction_config (dict): Configuration parameters for extraction

 Returns:
 dict: Extraction results including ArgDown content and validation info

 Raises:
 Exception: For any errors during extraction
 """
 print(f"Starting extraction from {extraction_config['source_path']}")
 start_time = time.time()

 try:
 # Process the document
 results = process_source_document(
 extraction_config["source_path"],
 provider_name=extraction_config["provider"]
)

 # Print success message
 elapsed_time = time.time() - start_time
 print(f"Extraction completed in {elapsed_time:.2f} seconds")
 print(f"Extracted {results['validation']['stats']['node_count']} nodes with "
 f"{results['validation']['stats']['relationship_count']} relationships")

 # Print any warnings
 if results['validation']['warnings']:
```



```

 print("\nWarnings:")
 for warning in results['validation']['warnings']:
 print(f"- {warning}")

 return results

except Exception as e:
 print(f"Error during extraction: {str(e)}")
 raise

Usage example:
extraction_results = execute_extraction(extraction_config)

```

Starting extraction from example\_document.pdf

Error during extraction: PyPDF2 is required for PDF processing. Install it with: pip install

ImportError: PyPDF2 is required for PDF processing. Install it with: pip install PyPDF2

```

ModuleNotFoundError Traceback (most recent call last)
<ipython-input-17-fd592eb962ab> in process_source_document(file_path, provider_name)
 166 try:
--> 167 import PyPDF2
 168 with open(file_path, 'rb') as file:
ModuleNotFoundError: No module named 'PyPDF2'
During handling of the above exception, another exception occurred:
ImportError Traceback (most recent call last)
<ipython-input-19-27555067c1d2> in <cell line: 0>()
 59
 60 # Usage example:
--> 61 extraction_results = execute_extraction(extraction_config)

<ipython-input-19-27555067c1d2> in execute_extraction(extraction_config)
 35 try:
 36 # Process the document
--> 37 results = process_source_document(
 38 extraction_config["source_path"],
 39 provider_name=extraction_config["provider"]
<ipython-input-17-fd592eb962ab> in process_source_document(file_path, provider_name)
 172 text += page.extract_text() + "\n"
 173 except ImportError:
--> 174 raise ImportError("PyPDF2 is required for PDF processing. Install it with
 175 elif file_path.endswith(".txt"):

```

```
176 with open(file_path, 'r') as file:
ImportError: PyPDF2 is required for PDF processing. Install it with: pip install PyPDF2
```

---

NOTE: If your import is failing due to a missing package, you can manually install dependencies using either !pip or !apt. To view examples of installing some common dependencies, click the "Open Examples" button below.

---

## 1.5 Save ArgDown Extraction Response

1. Save and log API return
2. Save ArgDown.md file for further Processing

```
@title 1.5 --- Save ArgDown Extraction Response ---

"""
BLOCK PURPOSE: Saves the extracted ArgDown content to files for further processing.

This function handles saving the extraction results:
1. Creates an output directory if it doesn't exist
2. Saves the extracted ArgDown content with a timestamp in the filename
3. Saves accompanying metadata in a JSON file
4. Saves a copy at a standard location for the next steps in the pipeline
5. Provides a preview of the extracted content

The saved files serve as inputs for the next stage of the pipeline where
probability information will be added to create BayesDown.

DEPENDENCIES: os module for directory operations
OUTPUTS: Saved ArgDown files and preview of extracted content
"""

def save_extraction_results(results, output_directory="./outputs"):
 """
 Save the extraction results to file

 Args:
 results (dict): Extraction results from execute_extraction
 output_directory (str): Directory to save results
```

```

Returns:
 str: Path to the saved ArgDown file
"""
Ensure output directory exists
import os
os.makedirs(output_directory, exist_ok=True)

Create base filename from source
import os.path
base_name = os.path.basename(results["source_path"]).split('.')[0]
timestamp = time.strftime("%Y%m%d-%H%M%S")
output_filename = f"{base_name}_argdown_{timestamp}.md"
output_path = os.path.join(output_directory, output_filename)

Save the results
save_argdown_extraction(results, output_path)

print(f"Saved ArgDown extraction to: {output_path}")
print(f"Metadata saved to: {output_path.replace('.md', '_metadata.json')}")

Also save to standard location for further processing
standard_path = os.path.join(output_directory, "ArgDown.md")
with open(standard_path, 'w') as f:
 f.write(results["argdown_content"])
print(f"Also saved to standard location: {standard_path}")

return output_path

Usage example:
output_path = save_extraction_results(extraction_results)

Preview the extracted ArgDown
from IPython.display import Markdown, display

Display the first 500 characters of the extracted ArgDown
preview = extraction_results["argdown_content"][:500] + "..." if len(extraction_results["argdown_content"]) > 500 else extraction_results["argdown_content"]
display(Markdown(f"## Extracted ArgDown Preview\n\n```\n{preview}\n```"))

```

NameError: name 'extraction\_results' is not defined

---

```

NameError Traceback (most recent call last)
<ipython-input-20-84ee4ea64739> in <cell line: 0>()

```

```

55
56 # Usage example:
--> 57 output_path = save_extraction_results(extraction_results)
58
59 # Preview the extracted ArgDown
NameError: name 'extraction_results' is not defined

```

## 1.6 Review and Check ArgDown.md File

```
display(Markdown(md_content))
```

[Existential\_Catastrophe]: The destruction of humanity’s long-term potential due to AI systems we’ve lost control over. {"instantiations": ["existential\_catastrophe\_TRUE", "existential\_catastrophe\_FALSE"]} - [Human\_Disempowerment]: Permanent and collective disempowerment of humanity relative to AI systems. {"instantiations": ["human\_disempowerment\_TRUE", "human\_disempowerment\_FALSE"]} - [Scale\_Of\_Power\_Seeking]: Power-seeking by AI systems scaling to the point of permanently disempowering all of humanity. {"instantiations": ["scale\_of\_power\_seeking\_TRUE", "scale\_of\_power\_seeking\_FALSE"]} - [Misaligned\_Power\_Seeking]: Deployed AI systems seeking power in unintended and high-impact ways due to problems with their objectives. {"instantiations": ["misaligned\_power\_seeking\_TRUE", "misaligned\_power\_seeking\_FALSE"]} - [APS\_Systems]: AI systems with advanced capabilities, agentic planning, and strategic awareness. {"instantiations": ["aps\_systems\_TRUE", "aps\_systems\_FALSE"]} - [Advanced\_AI\_Capability]: AI systems that outperform humans on tasks that grant significant power in the world. {"instantiations": ["advanced\_ai\_capability\_TRUE", "advanced\_ai\_capability\_FALSE"]} - [Agentic\_Planning]: AI systems making and executing plans based on world models to achieve objectives. {"instantiations": ["agentic\_planning\_TRUE", "agentic\_planning\_FALSE"]} - [Strategic\_Awareness]: AI systems with models accurately representing power dynamics with humans. {"instantiations": ["strategic\_awareness\_TRUE", "strategic\_awareness\_FALSE"]} - [Difficulty\_Of\_Alignment]: It is harder to build aligned systems than misaligned systems that are attractive to deploy. {"instantiations": ["difficulty\_of\_alignment\_TRUE", "difficulty\_of\_alignment\_FALSE"]} - [Instrumental\_Convergence]: AI systems with misaligned objectives tend to seek power as an instrumental goal. {"instantiations": ["instrumental\_convergence\_TRUE", "instrumental\_convergence\_FALSE"]} - [Problems\_With\_Proxies]: Optimizing for proxy objectives breaks correlations with intended goals. {"instantiations": ["problems\_with\_proxies\_TRUE", "problems\_with\_proxies\_FALSE"]} - [Problems\_With\_Search]: Search processes can yield systems pursuing different objectives than intended. {"instantiations": ["problems\_with\_search\_TRUE", "problems\_with\_search\_FALSE"]} - [Deployment\_Decisions]: Decisions to deploy potentially misaligned AI systems. {"instantiations": ["deployment\_decisions\_DEPLOY", "deployment\_decisions\_WITHHOLD"]} - [Incentives\_To\_Build\_APS]: Strong incentives to build

and deploy APS systems. {"instantiations": ["incentives\_to\_build\_aps\_STRONG", "incentives\_to\_build\_aps\_WEAK"]} - [Usefulness\_Of\_APS]: APS systems are very useful for many valuable tasks. {"instantiations": ["usefulness\_of\_aps\_HIGH", "usefulness\_of\_aps\_LOW"]} - [Competitive\_Dynamics]: Competitive pressures between AI developers. {"instantiations": ["competitive\_dynamics\_STRONG", "competitive\_dynamics\_WEAK"]} - [Deception\_By\_AI]: AI systems deceiving humans about their true objectives. {"instantiations": ["deception\_by\_ai\_TRUE", "deception\_by\_ai\_FALSE"]} - [Corrective\_Feedback]: Human society implementing corrections after observing problems. {"instantiations": ["corrective\_feedback\_EFFECTIVE", "corrective\_feedback\_INEFFECTIVE"]} - [Warning\_Shots]: Observable failures in weaker systems before catastrophic risks. {"instantiations": ["warning\_shots\_OBSERVED", "warning\_shots\_UNOBSERVED"]} - [Rapid\_Capability\_Escalation]: AI capabilities escalating very rapidly, allowing little time for correction. {"instantiations": ["rapid\_capability\_escalation\_TRUE", "rapid\_capability\_escalation\_FALSE"]} [Barriers\_To\_Understanding]: Difficulty in understanding the internal workings of advanced AI systems. {"instantiations": ["barriers\_to\_understanding\_HIGH", "barriers\_to\_understanding\_LOW"]} - [Misaligned\_Power\_Seeking]: Deployed AI systems seeking power in unintended and high-impact ways due to problems with their objectives. {"instantiations": ["misaligned\_power\_seeking\_TRUE", "misaligned\_power\_seeking\_FALSE"]} [Adversarial\_Dynamics]: Potentially adversarial relationships between humans and power-seeking AI. {"instantiations": ["adversarial\_dynamics\_TRUE", "adversarial\_dynamics\_FALSE"]} - [Misaligned\_Power\_Seeking]: Deployed AI systems seeking power in unintended and high-impact ways due to problems with their objectives. {"instantiations": ["misaligned\_power\_seeking\_TRUE", "misaligned\_power\_seeking\_FALSE"]} [Stakes\_Of\_Error]: The escalating impact of mistakes with power-seeking AI systems. {"instantiations": ["stakes\_of\_error\_HIGH", "stakes\_of\_error\_LOW"]} - [Misaligned\_Power\_Seeking]: Deployed AI systems seeking power in unintended and high-impact ways due to problems with their objectives. {"instantiations": ["misaligned\_power\_seeking\_TRUE", "misaligned\_power\_seeking\_FALSE"]}

## 1.6.2 Check the Graph Structure with the ArgDown Sandbox Online

Copy and paste the BayesDown formatted ... in the ArgDown Sandbox below to quickly verify that the network renders correctly.

```
@title 1.6.2 --- ArgDown Online Sandbox ---

from IPython.display import IFrame

IFrame(src="https://argdown.org/sandbox/map/", width="100%", height="600px")
```

```
<IPython.lib.display.IFrame at 0x7d3db5e18a10>
```

## 1.7 Extract ArgDown Graph Information as DataFrame

Extract:

- Nodes (Variable\_Title)
- Edges (Parents)
- Instantiations
- Description

Implementation nodes: - One function for ArgDown and BayesDown extraction, but: - IF YOU ONLY WANT ARGDOWN EXTRACTION: USE ARGUMENT IN FUNCTION CALL “parse\_markdown\_hierarchy(markdown\_text, ArgDown = True)” - so if you set ArgDown = True, it gives you only instantiations, no probabilities.

```
@title 1.7 --- Parsing ArgDown & BayesDown (.md to .csv) ---

"""
BLOCK PURPOSE: Provides the core parsing functionality for transforming ArgDown and BayesDown
text representations into structured DataFrame format for further processing.

This block implements the critical extraction pipeline described in the AMTAIR project
(see PY_TechnicalImplementation) that converts argument structures into Bayesian networks.
The function can handle both basic ArgDown (structure-only) and BayesDown (with probabilities).

Key steps in the parsing process:
1. Remove comments from the markdown text
2. Extract titles, descriptions, and indentation levels
3. Establish parent-child relationships based on indentation
4. Convert the structured information into a DataFrame
5. Add derived columns for network analysis

DEPENDENCIES: pandas, re, json libraries
INPUTS: Markdown text in ArgDown/BayesDown format
OUTPUTS: Structured DataFrame with node information, relationships, and properties
"""

def parse_markdown_hierarchy_fixed(markdown_text, ArgDown=False):
 """
 Parse ArgDown or BayesDown format into a structured DataFrame with parent-child relationships.

 Args:
 markdown_text (str): Text in ArgDown or BayesDown format
 ArgDown (bool): If True, extracts only structure without probabilities
 """
```

If False, extracts both structure and probability information

Returns:

pandas.DataFrame: Structured data with node information, relationships, and attributes

"""

# PHASE 1: Clean and prepare the text

clean\_text = remove\_comments(markdown\_text)

# PHASE 2: Extract basic information about nodes

titles\_info = extract\_titles\_info(clean\_text)

# PHASE 3: Determine the hierarchical relationships

titles\_with\_relations = establish\_relationships\_fixed(titles\_info, clean\_text)

# PHASE 4: Convert to structured DataFrame format

df = convert\_to\_dataframe(titles\_with\_relations, ArgDown)

# PHASE 5: Add derived columns for analysis

df = add\_no\_parent\_no\_child\_columns\_to\_df(df)

df = add\_parents\_instantiation\_columns\_to\_df(df)

return df

def remove\_comments(markdown\_text):

"""

Remove comment blocks from markdown text using regex pattern matching.

Args:

markdown\_text (str): Text containing potential comment blocks

Returns:

str: Text with comment blocks removed

"""

# Remove anything between /\* and \*/ using regex

return re.sub(r'/\\*.\*?\\*/', '', markdown\_text, flags=re.DOTALL)

def extract\_titles\_info(text):

"""

Extract titles with their descriptions and indentation levels from markdown text.

Args:

text (str): Cleaned markdown text

```

Returns:
 dict: Dictionary with titles as keys and dictionaries of attributes as values
 """
lines = text.split('\n')
titles_info = {}

for line in lines:
 # Skip empty lines
 if not line.strip():
 continue

 # Extract title within square or angle brackets
 title_match = re.search(r'<\[(.+?) >\]', line)
 if not title_match:
 continue

 title = title_match.group(1)

 # Extract description and metadata
 title_pattern_in_line = r'<\[' + re.escape(title) + r'>\]:'
 description_match = re.search(title_pattern_in_line + r'\s*(.*)', line)

 if description_match:
 full_text = description_match.group(1).strip()

 # Split description and metadata at the first "{"
 if "{" in full_text:
 split_index = full_text.find("{")
 description = full_text[:split_index].strip()
 metadata = full_text[split_index:].strip()
 else:
 # Keep the entire description and no metadata
 description = full_text
 metadata = '' # Initialize as empty string
 else:
 description = ''
 metadata = '' # Ensure metadata is initialized

 # Calculate indentation level based on spaces before + or - symbol
 indentation = 0
 if '+' in line:
 symbol_index = line.find('+')

```



```

 # Count spaces before the '+' symbol
 i = symbol_index - 1
 while i >= 0 and line[i] == ' ':
 indentation += 1
 i -= 1
 elif '-' in line:
 symbol_index = line.find('-')
 # Count spaces before the '-' symbol
 i = symbol_index - 1
 while i >= 0 and line[i] == ' ':
 indentation += 1
 i -= 1

If neither symbol exists, indentation remains 0

if title in titles_info:
 # Only update description if it's currently empty and we found a new one
 if not titles_info[title]['description'] and description:
 titles_info[title]['description'] = description

 # Store all indentation levels for this title
 titles_info[title]['indentation_levels'].append(indentation)

 # Keep max indentation for backward compatibility
 if indentation > titles_info[title]['indentation']:
 titles_info[title]['indentation'] = indentation

 # Do NOT update metadata here - keep the original metadata
else:
 # First time seeing this title, create a new entry
 titles_info[title] = {
 'description': description,
 'indentation': indentation,
 'indentation_levels': [indentation], # Initialize with first indentation level
 'parents': [],
 'children': [],
 'line': None,
 'line_numbers': [], # Initialize an empty list for all occurrences
 'metadata': metadata # Set metadata explicitly from what we found
 }

return titles_info

```

```

def establish_relationships_fixed(titles_info, text):
 """
 Establish parent-child relationships between titles using BayesDown indentation rules.

 In BayesDown syntax:
 - More indented nodes (with + symbol) are PARENTS of less indented nodes
 - The relationship reads as "Effect is caused by Cause" (Effect + Cause)
 - This aligns with how Bayesian networks represent causality

 Args:
 titles_info (dict): Dictionary with information about titles
 text (str): Original markdown text (for identifying line numbers)

 Returns:
 dict: Updated dictionary with parent-child relationships
 """
 lines = text.split('\n')

 # Dictionary to store line numbers for each title occurrence
 title_occurrences = {}

 # Record line number for each title (including multiple occurrences)
 line_number = 0
 for line in lines:
 if not line.strip():
 line_number += 1
 continue

 title_match = re.search(r'<\[(.+?) >\]', line)
 if not title_match:
 line_number += 1
 continue

 title = title_match.group(1)

 # Store all occurrences of each title with their line numbers
 if title not in title_occurrences:
 title_occurrences[title] = []
 title_occurrences[title].append(line_number)

 # Store all line numbers where this title appears
 if 'line_numbers' not in titles_info[title]:

```

```

 titles_info[title]['line_numbers'] = []
 titles_info[title]['line_numbers'].append(line_number)

 # For backward compatibility, keep the first occurrence in 'line'
 if titles_info[title]['line'] is None:
 titles_info[title]['line'] = line_number

 line_number += 1

Create an ordered list of all title occurrences with their line numbers
all_occurrences = []
for title, occurrences in title_occurrences.items():
 for line_num in occurrences:
 all_occurrences.append((title, line_num))

Sort occurrences by line number
all_occurrences.sort(key=lambda x: x[1])

Get indentation for each occurrence
occurrence_indents = {}
for title, line_num in all_occurrences:
 for line in lines[line_num:line_num+1]: # Only check the current line
 indent = 0
 if '+' in line:
 symbol_index = line.find('+')
 # Count spaces before the '+' symbol
 j = symbol_index - 1
 while j >= 0 and line[j] == ' ':
 indent += 1
 j -= 1
 elif '-' in line:
 symbol_index = line.find('-')
 # Count spaces before the '-' symbol
 j = symbol_index - 1
 while j >= 0 and line[j] == ' ':
 indent += 1
 j -= 1
 occurrence_indents[(title, line_num)] = indent

Enhanced backward pass for correct parent-child relationships
for i, (title, line_num) in enumerate(all_occurrences):
 current_indent = occurrence_indents[(title, line_num)]

```

```

 # Skip root nodes (indentation 0) for processing
 if current_indent == 0:
 continue

 # Look for the immediately preceding node with lower indentation
 j = i - 1
 while j >= 0:
 prev_title, prev_line = all_occurrences[j]
 prev_indent = occurrence_indents[(prev_title, prev_line)]

 # If we find a node with less indentation, it's a child of current node
 if prev_indent < current_indent:
 # In BayesDown: More indented node is a parent (cause) of less indented node
 if title not in titles_info[prev_title]['parents']:
 titles_info[prev_title]['parents'].append(title)
 if prev_title not in titles_info[title]['children']:
 titles_info[title]['children'].append(prev_title)

 # Only need to find the immediate child (closest preceding node with lower indentation)
 break

 j -= 1

 return titles_info

def convert_to_dataframe(titles_info, ArgDown):
 """
 Convert the titles information dictionary to a pandas DataFrame.

 Args:
 titles_info (dict): Dictionary with information about titles
 ArgDown (bool): If True, extract only structural information without probabilities

 Returns:
 pandas.DataFrame: Structured data with node information and relationships
 """
 if ArgDown == True:
 # For ArgDown, exclude probability columns
 df = pd.DataFrame(columns=['Title', 'Description', 'line', 'line_numbers', 'indentation',
 'indentation_levels', 'Parents', 'Children', 'instantiations'])
 else:
 # For BayesDown, include probability columns

```

```

df = pd.DataFrame(columns=['Title', 'Description', 'line', 'line_numbers', 'indentation',
 'indentation_levels', 'Parents', 'Children', 'instantiations',
 'priors', 'posteriors'])

for title, info in titles_info.items():
 # Parse the metadata JSON string into a Python dictionary
 if 'metadata' in info and info['metadata']:
 try:
 # Only try to parse if metadata is not empty
 if info['metadata'].strip():
 jsonMetadata = json.loads(info['metadata'])
 if ArgDown == True:
 # Create the row dictionary with instantiations as metadata only, no
 row = {
 'Title': title,
 'Description': info.get('description', ''),
 'line': info.get('line', ''),
 'line_numbers': info.get('line_numbers', []),
 'indentation': info.get('indentation', ''),
 'indentation_levels': info.get('indentation_levels', []),
 'Parents': info.get('parents', []),
 'Children': info.get('children', []),
 # Extract specific metadata fields, defaulting to empty if not present
 'instantiations': jsonMetadata.get('instantiations', []),
 }
 else:
 # Create dict with probabilities for BayesDown
 row = {
 'Title': title,
 'Description': info.get('description', ''),
 'line': info.get('line', ''),
 'line_numbers': info.get('line_numbers', []),
 'indentation': info.get('indentation', ''),
 'indentation_levels': info.get('indentation_levels', []),
 'Parents': info.get('parents', []),
 'Children': info.get('children', []),
 # Extract specific metadata fields, defaulting to empty if not present
 'instantiations': jsonMetadata.get('instantiations', []),
 'priors': jsonMetadata.get('priors', {}),
 'posteriors': jsonMetadata.get('posteriors', {})
 }
 else:
 else:

```

```

 # Empty metadata case
 row = {
 'Title': title,
 'Description': info.get('description', ''),
 'line': info.get('line', ''),
 'line_numbers': info.get('line_numbers', []),
 'indentation': info.get('indentation', ''),
 'indentation_levels': info.get('indentation_levels', []),
 'Parents': info.get('parents', []),
 'Children': info.get('children', []),
 'instantiations': [],
 'priors': {},
 'posteriors': {}
 }
 except json.JSONDecodeError:
 # Handle case where metadata isn't valid JSON
 row = {
 'Title': title,
 'Description': info.get('description', ''),
 'line': info.get('line', ''),
 'line_numbers': info.get('line_numbers', []),
 'indentation': info.get('indentation', ''),
 'indentation_levels': info.get('indentation_levels', []),
 'Parents': info.get('parents', []),
 'Children': info.get('children', []),
 'instantiations': [],
 'priors': {},
 'posteriors': {}
 }
 else:
 # Handle case where metadata field doesn't exist or is empty
 row = {
 'Title': title,
 'Description': info.get('description', ''),
 'line': info.get('line', ''),
 'line_numbers': info.get('line_numbers', []),
 'indentation': info.get('indentation', ''),
 'indentation_levels': info.get('indentation_levels', []),
 'Parents': info.get('parents', []),
 'Children': info.get('children', []),
 'instantiations': [],
 'priors': {},

```

```

 'posteriors': {}
 }

 # Add the row to the DataFrame
 df.loc[len(df)] = row

 return df

def add_no_parent_no_child_columns_to_df(dataframe):
 """
 Add No_Parent and No_Children boolean columns to the DataFrame to identify root and leaf

 Args:
 dataframe (pandas.DataFrame): The DataFrame to enhance

 Returns:
 pandas.DataFrame: Enhanced DataFrame with additional boolean columns
 """
 no_parent = []
 no_children = []

 for _, row in dataframe.iterrows():
 no_parent.append(not row['Parents']) # True if Parents list is empty
 no_children.append(not row['Children']) # True if Children list is empty

 dataframe['No_Parent'] = no_parent
 dataframe['No_Children'] = no_children

 return dataframe

def add_parents_instantiation_columns_to_df(dataframe):
 """
 Add all possible instantiations of parents as a list of lists column to the DataFrame.
 This is crucial for generating conditional probability tables.

 Args:
 dataframe (pandas.DataFrame): The DataFrame to enhance

 Returns:
 pandas.DataFrame: Enhanced DataFrame with parent_instantiations column
 """
 # Create a new column to store parent instantiations

```

```

parent_instantiations = []

Iterate through each row in the dataframe
for _, row in dataframe.iterrows():
 parents = row['Parents']
 parent_insts = []

 # For each parent, find its instantiations and add to the list
 for parent in parents:
 # Find the row where Title matches the parent
 parent_row = dataframe[dataframe['Title'] == parent]

 # If parent found in the dataframe
 if not parent_row.empty:
 # Get the instantiations of this parent
 parent_instantiation = parent_row['instantiations'].iloc[0]
 parent_insts.append(parent_instantiation)

 # Add the list of parent instantiations to our new column
 parent_instantiations.append(parent_insts)

Add the new column to the dataframe
dataframe['parent_instantiations'] = parent_instantiations

return dataframe

```

```

example use case:
ex_csv = parse_markdown_hierarchy_fixed(md_content, ArgDown = True)
ex_csv

```

	Title	Description	line	line_number
0	Existential_Catastrophe	The destruction of humanity's long-term potent...	0	[0]
1	Human_Disempowerment	Permanent and collective disempowerment of hum...	1	[1]
2	Scale_Of_Power_Seeking	Power-seeking by AI systems scaling to the poi...	2	[2]
3	Misaligned_Power_Seeking	Deployed AI systems seeking power in unintende...	3	[3, 21, 23, 25]
4	APS_Systems	AI systems with advanced capabilities, agentic...	4	[4]
5	Advanced_AI_Capability	AI systems that outperform humans on tasks tha...	5	[5]
6	Agentic_Planning	AI systems making and executing plans based on...	6	[6]
7	Strategic_Awareness	AI systems with models accurately representing...	7	[7]
8	Difficulty_Of_Alignment	It is harder to build aligned systems than mis...	8	[8]
9	Instrumental_Convergence	AI systems with misaligned objectives tend to ...	9	[9]



	Title	Description	line	line_number
10	Problems_With_Proxies	Optimizing for proxy objectives breaks correla...	10	[10]
11	Problems_With_Search	Search processes can yield systems pursuing di...	11	[11]
12	Deployment_Decisions	Decisions to deploy potentially misaligned AI ...	12	[12]
13	Incentives_To_Build_APS	Strong incentives to build and deploy APS syst...	13	[13]
14	Usefulness_Of_APS	APS systems are very useful for many valuable ...	14	[14]
15	Competitive_Dynamics	Competitive pressures between AI developers.	15	[15]
16	Deception_By_AI	AI systems deceiving humans about their true o...	16	[16]
17	Corrective_Feedback	Human society implementing corrections after o...	17	[17]
18	Warning_Shots	Observable failures in weaker systems before c...	18	[18]
19	Rapid_Capability_Escalation	AI capabilities escalating very rapidly, allow...	19	[19]
20	Barriers_To_Understanding	Difficulty in understanding the internal worki...	20	[20]
21	Adversarial_Dynamics	Potentially adversarial relationships between ...	22	[22]
22	Stakes_Of_Error	The escalating impact of mistakes with power-s...	24	[24]

## 1.8 Store ArgDown Information as 'ArgDown.csv' file

```
Assuming 'md_content' holds the markdown text
Store the results of running the function parse_markdown_hierarchy(md_content, ArgDown = T
result_df = parse_markdown_hierarchy_fixed(md_content, ArgDown = True)

Save to CSV
result_df.to_csv('ArgDown.csv', index=False)

Test if 'ArgDown.csv' has been saved correctly with the correct information
Load the data from the CSV file
argdown_df = pd.read_csv('ArgDown.csv')

Display the DataFrame
print(argdown_df)
```

```

 Title \
0 Existential_Catastrophe
1 Human_Disempowerment
2 Scale_Of_Power_Seeking
3 Misaligned_Power_Seeking
4 APS_Systems
5 Advanced_AI_Capability
6 Agentic_Planning
```

```

7 Strategic_Awareness
8 Difficulty_Of_Alignment
9 Instrumental_Convergence
10 Problems_With_Proxies
11 Problems_With_Search
12 Deployment_Decisions
13 Incentives_To_Build_APS
14 Usefulness_Of_APS
15 Competitive_Dynamics
16 Deception_By_AI
17 Corrective_Feedback
18 Warning_Shots
19 Rapid_Capability_Escalation
20 Barriers_To_Understanding
21 Adversarial_Dynamics
22 Stakes_Of_Error

```

	Description	line	line_numbers \
0	The destruction of humanity's long-term potent...	0	[0]
1	Permanent and collective disempowerment of hum...	1	[1]
2	Power-seeking by AI systems scaling to the poi...	2	[2]
3	Deployed AI systems seeking power in unintende...	3	[3, 21, 23, 25]
4	AI systems with advanced capabilities, agentic...	4	[4]
5	AI systems that outperform humans on tasks tha...	5	[5]
6	AI systems making and executing plans based on...	6	[6]
7	AI systems with models accurately representing...	7	[7]
8	It is harder to build aligned systems than mis...	8	[8]
9	AI systems with misaligned objectives tend to ...	9	[9]
10	Optimizing for proxy objectives breaks correla...	10	[10]
11	Search processes can yield systems pursuing di...	11	[11]
12	Decisions to deploy potentially misaligned AI ...	12	[12]
13	Strong incentives to build and deploy APS syst...	13	[13]
14	APS systems are very useful for many valuable ...	14	[14]
15	Competitive pressures between AI developers.	15	[15]
16	AI systems deceiving humans about their true o...	16	[16]
17	Human society implementing corrections after o...	17	[17]
18	Observable failures in weaker systems before c...	18	[18]
19	AI capabilities escalating very rapidly, allow...	19	[19]
20	Difficulty in understanding the internal worki...	20	[20]
21	Potentially adversarial relationships between ...	22	[22]
22	The escalating impact of mistakes with power-s...	24	[24]

```

indentation indentation_levels \

```

0	0	[0]
1	0	[0]
2	4	[4]
3	8	[8, 0, 0, 0]
4	12	[12]
5	16	[16]
6	16	[16]
7	16	[16]
8	12	[12]
9	16	[16]
10	16	[16]
11	16	[16]
12	12	[12]
13	16	[16]
14	20	[20]
15	20	[20]
16	16	[16]
17	8	[8]
18	12	[12]
19	12	[12]
20	0	[0]
21	0	[0]
22	0	[0]

	Parents	\
0		[]
1		['Scale_Of_Power_Seeking']
2		['Misaligned_Power_Seeking', 'Corrective_Feedb...]
3		['APS_Systems', 'Difficulty_Of_Alignment', 'De...]
4		['Advanced_AI_Capability', 'Agentic_Planning',...]
5		[]
6		[]
7		[]
8		['Instrumental_Convergence', 'Problems_With_Pr...]
9		[]
10		[]
11		[]
12		['Incentives_To_Build_APS', 'Deception_By_AI']
13		['Usefulness_Of_APS', 'Competitive_Dynamics']
14		[]
15		[]
16		[]
17		['Warning_Shots', 'Rapid_Capability_Escalation']

18	[]
19	[]
20	[]
21	[]
22	[]

	Children \
0	[]
1	[]
2	['Human_Disempowerment']
3	['Scale_Of_Power_Seeking']
4	['Misaligned_Power_Seeking']
5	['APS_Systems']
6	['APS_Systems']
7	['APS_Systems']
8	['Misaligned_Power_Seeking']
9	['Difficulty_Of_Alignment']
10	['Difficulty_Of_Alignment']
11	['Difficulty_Of_Alignment']
12	['Misaligned_Power_Seeking']
13	['Deployment_Decisions']
14	['Incentives_To_Build_APS']
15	['Incentives_To_Build_APS']
16	['Deployment_Decisions']
17	['Scale_Of_Power_Seeking']
18	['Corrective_Feedback']
19	['Corrective_Feedback']
20	[]
21	[]
22	[]

	instantiations	No_Parent	No_Children \
0	['existential_catastrophe_TRUE', 'existential_...	True	True
1	['human_disempowerment_TRUE', 'human_disempowe...	False	True
2	['scale_of_power_seeking_TRUE', 'scale_of_powe...	False	False
3	['misaligned_power_seeking_TRUE', 'misaligned_...	False	False
4	['aps_systems_TRUE', 'aps_systems_FALSE']	False	False
5	['advanced_ai_capability_TRUE', 'advanced_ai_c...	True	False
6	['agentic_planning_TRUE', 'agentic_planning_FA...	True	False
7	['strategic_awareness_TRUE', 'strategic_aware...	True	False
8	['difficulty_of_alignment_TRUE', 'difficulty_o...	False	False
9	['instrumental_convergence_TRUE', 'instrumenta...	True	False
10	['problems_with_proxies_TRUE', 'problems_with_...	True	False

11	['problems_with_search_TRUE', 'problems_with_s...	True	False
12	['deployment_decisions_DEPLOY', 'deployment_de...	False	False
13	['incentives_to_build_aps_STRONG', 'incentives...	False	False
14	['usefulness_of_aps_HIGH', 'usefulness_of_aps_...	True	False
15	['competitive_dynamics_STRONG', 'competitive_d...	True	False
16	['deception_by_ai_TRUE', 'deception_by_ai_FALSE']	True	False
17	['corrective_feedback_EFFECTIVE', 'corrective_...	False	False
18	['warning_shots_OBSERVED', 'warning_shots_UNOB...	True	False
19	['rapid_capability_escalation_TRUE', 'rapid_ca...	True	False
20	['barriers_to_understanding_HIGH', 'barriers_t...	True	True
21	['adversarial_dynamics_TRUE', 'adversarial_dyn...	True	True
22	['stakes_of_error_HIGH', 'stakes_of_error_LOW']	True	True

```

 parent_instantiations
0 []
1 [['scale_of_power_seeking_TRUE', 'scale_of_pow...
2 [['misaligned_power_seeking_TRUE', 'misaligned...
3 [['aps_systems_TRUE', 'aps_systems_FALSE'], ['...
4 [['advanced_ai_capability_TRUE', 'advanced_ai_...
5 []
6 []
7 []
8 [['instrumental_convergence_TRUE', 'instrument...
9 []
10 []
11 []
12 [['incentives_to_build_aps_STRONG', 'incentive...
13 [['usefulness_of_aps_HIGH', 'usefulness_of_aps...
14 []
15 []
16 []
17 [['warning_shots_OBSERVED', 'warning_shots_UNO...
18 []
19 []
20 []
21 []
22 []

```

## 2.0 Probability Extractions: ArgDown (.csv) to BayesDown (.md + plugin JSON syntax)

### 2. ArgDown to BayesDown: Adding Probability Information

#### Process Overview

This section implements the second major stage of the AMTAIR pipeline: enhancing the structured argument representation (ArgDown) with probability information to create BayesDown.

BayesDown extends ArgDown by adding: 1. Prior probabilities for each variable (unconditional beliefs) 2. Conditional probabilities representing the relationships between variables 3. The full parameter specification needed for a Bayesian network

The process follows these steps: 1. Generate probability questions for each node and its relationships 2. Create a BayesDown template with placeholders for these probabilities 3. Answer the probability questions (manually or via LLM) 4. Substitute the answers into the BayesDown representation

This enhanced representation contains all the information needed to construct a formal Bayesian network, enabling probabilistic reasoning and policy evaluation.

#### What is BayesDown?

BayesDown maintains the ArgDown structure but adds probability metadata:

```
[Node]: Description. {
 "instantiations": ["node_TRUE", "node_FALSE"],
 "priors": { "p(node_TRUE)": "0.7", "p(node_FALSE)": "0.3" },
 "posteriors": { "p(node_TRUE|parent_TRUE)": "0.9", "p(node_TRUE|parent_FALSE)": "0.4" }
}
```

The result is a hybrid representation that preserves the narrative structure of arguments while adding the mathematical precision of Bayesian networks.

#### 2.1 Probability Extraction Questions — ‘ArgDown.csv’ to ‘ArgDown\_WithQuestions.csv’

```

@title 2.1 --- Probability Extraction Questions Generation ---

"""
BLOCK PURPOSE: Generates probability questions for ArgDown nodes to prepare for BayesDown con

This block implements a key step in the pipeline where structure (from ArgDown) is prepared
for probability integration (to create BayesDown). It:

1. Processes a CSV file containing ArgDown structure
2. For each node, generates appropriate probability questions:
 - Prior probability questions for all nodes
 - Conditional probability questions for nodes with parents
3. Creates a new CSV file with these questions ready for the next stage

The generated questions serve as placeholders that will be answered in the
probability extraction phase to complete the Bayesian network.

DEPENDENCIES: pandas, json, itertools libraries
INPUTS: ArgDown CSV file
OUTPUTS: Enhanced CSV with probability questions for each node
"""

import pandas as pd
import re
import json
import itertools
from IPython.display import Markdown, display

def parse_instantiations(instantiations_str):
 """
 Parse instantiations from string or list format.
 Handles various input formats flexibly.

 Args:
 instantiations_str: Instantiations in string or list format

 Returns:
 list: Parsed instantiations as a list
 """
 if pd.isna(instantiations_str) or instantiations_str == '':

```

```

 return []

 if isinstance(instantiations_str, list):
 return instantiations_str

 try:
 # Try to parse as JSON
 return json.loads(instantiations_str)
 except:
 # Try to parse as string list
 if isinstance(instantiations_str, str):
 # Remove brackets and split by comma
 clean_str = instantiations_str.strip('[]"\'')
 if not clean_str:
 return []
 return [s.strip(' "') for s in clean_str.split(',') if s.strip()]

 return []

def parse_parents(parents_str):
 """
 Parse parents from string or list format.
 Handles various input formats flexibly.

 Args:
 parents_str: Parents in string or list format

 Returns:
 list: Parsed parents as a list
 """
 if pd.isna(parents_str) or parents_str == '':
 return []

 if isinstance(parents_str, list):
 return parents_str

 try:
 # Try to parse as JSON
 return json.loads(parents_str)
 except:
 # Try to parse as string list
 if isinstance(parents_str, str):

```



```

 # Remove brackets and split by comma
 clean_str = parents_str.strip('["\']')
 if not clean_str:
 return []
 return [s.strip(' "') for s in clean_str.split(',') if s.strip()]

 return []

def get_parent_instantiations(parent, df):
 """
 Get the instantiations for a parent node from the DataFrame.
 Returns default instantiations if not found.

 Args:
 parent (str): Parent node name
 df (DataFrame): DataFrame containing node information

 Returns:
 list: Instantiations for the parent node
 """
 parent_row = df[df['Title'] == parent]
 if parent_row.empty:
 return [f"{parent}_TRUE", f"{parent}_FALSE"]

 instantiations = parse_instantiations(parent_row.iloc[0]['instantiations'])
 if not instantiations:
 return [f"{parent}_TRUE", f"{parent}_FALSE"]

 return instantiations

def generate_instantiation_questions(title, instantiation, parents, df):
 """
 Generate questions for a specific instantiation of a node.

 Args:
 title (str): The title of the node
 instantiation (str): The specific instantiation (e.g., "title_TRUE")
 parents (list): List of parent nodes
 df (DataFrame): The full DataFrame for looking up parent instantiations

 Returns:
 dict: Dictionary mapping questions to estimate keys
 """

```

```

"""
questions = {}

Always generate a prior probability question, regardless of parents
prior_question = f"What is the probability for {title}={instantiation}?"
questions[prior_question] = 'prior' # Question is the key, 'prior' is the value

If no parents, return only the prior question
if not parents:
 return questions

For nodes with parents, generate conditional probability questions
Get all combinations of parent instantiations
parent_instantiations = []
for parent in parents:
 parent_insts = get_parent_instantiations(parent, df)
 parent_instantiations.append([(parent, inst) for inst in parent_insts])

Generate all combinations
all_combinations = list(itertools.product(*parent_instantiations))

Create conditional probability questions for each combination
and use questions as keys, estimate_i as values
for i, combination in enumerate(all_combinations):
 condition_str = ", ".join([f"{parent}={inst}" for parent, inst in combination])
 question = f"What is the probability for {title}={instantiation} if {condition_str}?"
 questions[question] = f'estimate_{i + 1}' # Question is the key, estimate_i is the value

return questions

def generate_argdown_with_questions(argdown_csv_path, output_csv_path):
 """
 Generate probability questions based on the ArgDown CSV file and save to a new CSV file.

 Args:
 argdown_csv_path (str): Path to the input ArgDown CSV file
 output_csv_path (str): Path to save the output CSV file with questions

 Returns:
 DataFrame: Enhanced DataFrame with probability questions

```

```

Raises:
 Exception: If CSV loading fails or required columns are missing
 """
print(f"Loading ArgDown CSV from {argdown_csv_path}...")

Load the ArgDown CSV file
try:
 df = pd.read_csv(argdown_csv_path)
 print(f"Successfully loaded CSV with {len(df)} rows.")
except Exception as e:
 raise Exception(f"Error loading ArgDown CSV: {e}")

Validate required columns
required_columns = ['Title', 'Parents', 'instantiations']
missing_columns = [col for col in required_columns if col not in df.columns]
if missing_columns:
 raise Exception(f"Missing required columns: {'', '.join(missing_columns)}")

Initialize columns for questions
df['Generate_Positive_Instantiation_Questions'] = None
df['Generate_Negative_Instantiation_Questions'] = None

print("Generating probability questions for each node...")

Process each row to generate questions
for idx, row in df.iterrows():
 title = row['Title']
 instantiations = parse_instantiations(row['instantiations'])
 parents = parse_parents(row['Parents'])

 if len(instantiations) < 2:
 # Default instantiations if not provided
 instantiations = [f"{title}_TRUE", f"{title}_FALSE"]

 # Generate positive instantiation questions
 positive_questions = generate_instantiation_questions(title, instantiations[0], parents)

 # Generate negative instantiation questions
 negative_questions = generate_instantiation_questions(title, instantiations[1], parents)

 # Update the DataFrame
 df.at[idx, 'Generate_Positive_Instantiation_Questions'] = json.dumps(positive_questions)

```

```

 df.at[idx, 'Generate_Negative_Instantiation_Questions'] = json.dumps(negative_questions)

 # Save the enhanced DataFrame
 df.to_csv(output_csv_path, index=False)
 print(f"Generated questions saved to {output_csv_path}")

 return df

Example usage:
df_with_questions = generate_argdown_with_questions("ArgDown.csv", "ArgDown_WithQuestions.csv")

```

Loading ArgDown CSV from ArgDown.csv...  
 Successfully loaded CSV with 23 rows.  
 Generating probability questions for each node...  
 Generated questions saved to ArgDown\_WithQuestions.csv

```

Load the data from the ArgDown_WithQuestions CSV file
argdown_with_questions_df = pd.read_csv('ArgDown_WithQuestions.csv')

Display the DataFrame
print(argdown_with_questions_df)
argdown_with_questions_df

```

	Title \
0	Existential_Catastrophe
1	Human_Disempowerment
2	Scale_Of_Power_Seeking
3	Misaligned_Power_Seeking
4	APS_Systems
5	Advanced_AI_Capability
6	Agentic_Planning
7	Strategic_Awareness
8	Difficulty_Of_Alignment
9	Instrumental_Convergence
10	Problems_With_Proxies
11	Problems_With_Search
12	Deployment_Decisions
13	Incentives_To_Build_APS
14	Usefulness_Of_APS
15	Competitive_Dynamics
16	Deception_By_AI

```

17 Corrective_Feedback
18 Warning_Shots
19 Rapid_Capability_Escalation
20 Barriers_To_Understanding
21 Adversarial_Dynamics
22 Stakes_Of_Error

```

	Description	line	line_numbers	\
0	The destruction of humanity's long-term potent...	0	[0]	
1	Permanent and collective disempowerment of hum...	1	[1]	
2	Power-seeking by AI systems scaling to the poi...	2	[2]	
3	Deployed AI systems seeking power in unintende...	3	[3, 21, 23, 25]	
4	AI systems with advanced capabilities, agentic...	4	[4]	
5	AI systems that outperform humans on tasks tha...	5	[5]	
6	AI systems making and executing plans based on...	6	[6]	
7	AI systems with models accurately representing...	7	[7]	
8	It is harder to build aligned systems than mis...	8	[8]	
9	AI systems with misaligned objectives tend to ...	9	[9]	
10	Optimizing for proxy objectives breaks correla...	10	[10]	
11	Search processes can yield systems pursuing di...	11	[11]	
12	Decisions to deploy potentially misaligned AI ...	12	[12]	
13	Strong incentives to build and deploy APS syst...	13	[13]	
14	APS systems are very useful for many valuable ...	14	[14]	
15	Competitive pressures between AI developers.	15	[15]	
16	AI systems deceiving humans about their true o...	16	[16]	
17	Human society implementing corrections after o...	17	[17]	
18	Observable failures in weaker systems before c...	18	[18]	
19	AI capabilities escalating very rapidly, allow...	19	[19]	
20	Difficulty in understanding the internal worki...	20	[20]	
21	Potentially adversarial relationships between ...	22	[22]	
22	The escalating impact of mistakes with power-s...	24	[24]	

	indentation	indentation_levels	\
0	0	[0]	
1	0	[0]	
2	4	[4]	
3	8	[8, 0, 0, 0]	
4	12	[12]	
5	16	[16]	
6	16	[16]	
7	16	[16]	
8	12	[12]	
9	16	[16]	

10	16	[16]
11	16	[16]
12	12	[12]
13	16	[16]
14	20	[20]
15	20	[20]
16	16	[16]
17	8	[8]
18	12	[12]
19	12	[12]
20	0	[0]
21	0	[0]
22	0	[0]

	Parents \
0	[]
1	['Scale_Of_Power_Seeking']
2	['Misaligned_Power_Seeking', 'Corrective_Feedb...]
3	['APS_Systems', 'Difficulty_Of_Alignment', 'De...]
4	['Advanced_AI_Capability', 'Agentic_Planning',...]
5	[]
6	[]
7	[]
8	['Instrumental_Convergence', 'Problems_With_Pr...]
9	[]
10	[]
11	[]
12	['Incentives_To_Build_APS', 'Deception_By_AI']
13	['Usefulness_Of_APS', 'Competitive_Dynamics']
14	[]
15	[]
16	[]
17	['Warning_Shots', 'Rapid_Capability_Escalation']
18	[]
19	[]
20	[]
21	[]
22	[]

	Children \
0	[]
1	[]
2	['Human_Disempowerment']

```

3 ['Scale_Of_Power_Seeking']
4 ['Misaligned_Power_Seeking']
5 ['APS_Systems']
6 ['APS_Systems']
7 ['APS_Systems']
8 ['Misaligned_Power_Seeking']
9 ['Difficulty_Of_Alignment']
10 ['Difficulty_Of_Alignment']
11 ['Difficulty_Of_Alignment']
12 ['Misaligned_Power_Seeking']
13 ['Deployment_Decisions']
14 ['Incentives_To_Build_APS']
15 ['Incentives_To_Build_APS']
16 ['Deployment_Decisions']
17 ['Scale_Of_Power_Seeking']
18 ['Corrective_Feedback']
19 ['Corrective_Feedback']
20
21
22

```

	instantiations	No_Parent	No_Children	\
0	['existential_catastrophe_TRUE', 'existential_...	True	True	
1	['human_disempowerment_TRUE', 'human_disempowe...	False	True	
2	['scale_of_power_seeking_TRUE', 'scale_of_powe...	False	False	
3	['misaligned_power_seeking_TRUE', 'misaligned_...	False	False	
4	['aps_systems_TRUE', 'aps_systems_FALSE']	False	False	
5	['advanced_ai_capability_TRUE', 'advanced_ai_c...	True	False	
6	['agentic_planning_TRUE', 'agentic_planning_FA...	True	False	
7	['strategic_awareness_TRUE', 'strategic_aware...	True	False	
8	['difficulty_of_alignment_TRUE', 'difficulty_o...	False	False	
9	['instrumental_convergence_TRUE', 'instrumenta...	True	False	
10	['problems_with_proxies_TRUE', 'problems_with_...	True	False	
11	['problems_with_search_TRUE', 'problems_with_s...	True	False	
12	['deployment_decisions_DEPLOY', 'deployment_de...	False	False	
13	['incentives_to_build_aps_STRONG', 'incentives...	False	False	
14	['usefulness_of_aps_HIGH', 'usefulness_of_aps_...	True	False	
15	['competitive_dynamics_STRONG', 'competitive_d...	True	False	
16	['deception_by_ai_TRUE', 'deception_by_ai_FALSE']	True	False	
17	['corrective_feedback_EFFECTIVE', 'corrective_...	False	False	
18	['warning_shots_OBSERVED', 'warning_shots_UNOB...	True	False	
19	['rapid_capability_escalation_TRUE', 'rapid_ca...	True	False	
20	['barriers_to_understanding_HIGH', 'barriers_t...	True	True	

21	['adversarial_dynamics_TRUE', 'adversarial_dyn...	True	True
22	['stakes_of_error_HIGH', 'stakes_of_error_LOW']	True	True

```

 parent_instantiations \
0 []
1 [['scale_of_power_seeking_TRUE', 'scale_of_pow...
2 [['misaligned_power_seeking_TRUE', 'misaligned...
3 [['aps_systems_TRUE', 'aps_systems_FALSE'], ['...
4 [['advanced_ai_capability_TRUE', 'advanced_ai_...
5 []
6 []
7 []
8 [['instrumental_convergence_TRUE', 'instrument...
9 []
10 []
11 []
12 [['incentives_to_build_aps_STRONG', 'incentive...
13 [['usefulness_of_aps_HIGH', 'usefulness_of_aps...
14 []
15 []
16 []
17 [['warning_shots_OBSERVED', 'warning_shots_UNO...
18 []
19 []
20 []
21 []
22 []

```

```

 Generate_Positive_Instantiation_Questions \
0 {"What is the probability for Existential_Cata...
1 {"What is the probability for Human_Disempower...
2 {"What is the probability for Scale_Of_Power_S...
3 {"What is the probability for Misaligned_Power...
4 {"What is the probability for APS_Systems=aps_...
5 {"What is the probability for Advanced_AI_Capa...
6 {"What is the probability for Agentic_Planning...
7 {"What is the probability for Strategic_Awaren...
8 {"What is the probability for Difficulty_Of_Al...
9 {"What is the probability for Instrumental_Con...
10 {"What is the probability for Problems_With_Pr...
11 {"What is the probability for Problems_With_Se...
12 {"What is the probability for Deployment_Decis...
13 {"What is the probability for Incentives_To_Bu...

```



```

14 {"What is the probability for Usefulness_Of_AP...
15 {"What is the probability for Competitive_Dyna...
16 {"What is the probability for Deception_By_AI=...
17 {"What is the probability for Corrective_Feedb...
18 {"What is the probability for Warning_Shots=wa...
19 {"What is the probability for Rapid_Capability...
20 {"What is the probability for Barriers_To_Unde...
21 {"What is the probability for Adversarial_Dyna...
22 {"What is the probability for Stakes_Of_Error=...

```

#### Generate\_Negative\_Instantiation\_Questions

```

0 {"What is the probability for Existential_Cata...
1 {"What is the probability for Human_Disempower...
2 {"What is the probability for Scale_Of_Power_S...
3 {"What is the probability for Misaligned_Power...
4 {"What is the probability for APS_Systems=aps_...
5 {"What is the probability for Advanced_AI_Capa...
6 {"What is the probability for Agentic_Planning...
7 {"What is the probability for Strategic_Awaren...
8 {"What is the probability for Difficulty_Of_Al...
9 {"What is the probability for Instrumental_Con...
10 {"What is the probability for Problems_With_Pr...
11 {"What is the probability for Problems_With_Se...
12 {"What is the probability for Deployment_Decis...
13 {"What is the probability for Incentives_To_Bu...
14 {"What is the probability for Usefulness_Of_AP...
15 {"What is the probability for Competitive_Dyna...
16 {"What is the probability for Deception_By_AI=...
17 {"What is the probability for Corrective_Feedb...
18 {"What is the probability for Warning_Shots=wa...
19 {"What is the probability for Rapid_Capability...
20 {"What is the probability for Barriers_To_Unde...
21 {"What is the probability for Adversarial_Dyna...
22 {"What is the probability for Stakes_Of_Error=...

```

	Title	Description	line	line_number
0	Existential_Catastrophe	The destruction of humanity's long-term potent...	0	[0]
1	Human_Disempowerment	Permanent and collective disempowerment of hum...	1	[1]
2	Scale_Of_Power_Seeking	Power-seeking by AI systems scaling to the poi...	2	[2]
3	Misaligned_Power_Seeking	Deployed AI systems seeking power in unintende...	3	[3, 21, 23, 25]
4	APS_Systems	AI systems with advanced capabilities, agentic...	4	[4]

	Title	Description	line	line_number
5	Advanced_AI_Capability	AI systems that outperform humans on tasks tha...	5	[5]
6	Agentic_Planning	AI systems making and executing plans based on...	6	[6]
7	Strategic_Awareness	AI systems with models accurately representing...	7	[7]
8	Difficulty_Of_Alignment	It is harder to build aligned systems than mis...	8	[8]
9	Instrumental_Convergence	AI systems with misaligned objectives tend to ...	9	[9]
10	Problems_With_Proxies	Optimizing for proxy objectives breaks correla...	10	[10]
11	Problems_With_Search	Search processes can yield systems pursuing di...	11	[11]
12	Deployment_Decisions	Decisions to deploy potentially misaligned AI ...	12	[12]
13	Incentives_To_Build_APS	Strong incentives to build and deploy APS syst...	13	[13]
14	Usefulness_Of_APS	APS systems are very useful for many valuable ...	14	[14]
15	Competitive_Dynamics	Competitive pressures between AI developers.	15	[15]
16	Deception_By_AI	AI systems deceiving humans about their true o...	16	[16]
17	Corrective_Feedback	Human society implementing corrections after o...	17	[17]
18	Warning_Shots	Observable failures in weaker systems before c...	18	[18]
19	Rapid_Capability_Escalation	AI capabilities escalating very rapidly, allow...	19	[19]
20	Barriers_To_Understanding	Difficulty in understanding the internal worki...	20	[20]
21	Adversarial_Dynamics	Potentially adversarial relationships between ...	22	[22]
22	Stakes_Of_Error	The escalating impact of mistakes with power-s...	24	[24]

## 2.2 ‘ArgDown\_WithQuestions.csv’ to ‘BayesDownQuestions.md’

### 2.2 Save BayesDown Extraction Questions as ‘BayesDownQuestions.md’

```
@title 2.2 --- BayesDown Questions Generation ---

"""
BLOCK PURPOSE: Transforms the ArgDown with questions into a BayesDown template with placeholders

This function creates a BayesDown representation with probability placeholders based on the c

1. Loads the CSV file with probability questions
2. Constructs a directed graph to represent the causal structure
3. Processes each node to create BayesDown syntax with probability placeholders
4. Optionally includes comments with the specific questions to be answered
5. Saves the result as a markdown file for the next stage of the pipeline

The output is a BayesDown template that can be used in the probability extraction phase, whee

DEPENDENCIES: networkx, pandas, json libraries
INPUTS: CSV file with ArgDown structure and probability questions
```

```

OUTPUTS: BayesDown markdown file with probability placeholders
"""

def extract_bayesdown_questions_fixed(argdown_with_questions_path, output_md_path, include_q
 """
 Generate BayesDown syntax from the ArgDown_WithQuestions CSV file with correct parent-child

 Args:
 argdown_with_questions_path (str): Path to the CSV file with probability questions
 output_md_path (str): Path to save the output BayesDown file
 include_questions_as_comments (bool, optional): Whether to include the original
 questions as comments. Defaults to True.

 Returns:
 str: The generated BayesDown content

 Raises:
 Exception: If CSV loading fails or required columns are missing
 """
 print(f"Loading CSV from {argdown_with_questions_path}...")

 # Load the CSV file
 try:
 df = pd.read_csv(argdown_with_questions_path)
 print(f"Successfully loaded CSV with {len(df)} rows.")
 except Exception as e:
 raise Exception(f"Error loading CSV: {e}")

 # Validate required columns
 required_columns = ['Title', 'Description', 'Parents', 'Children', 'instantiations']
 missing_columns = [col for col in required_columns if col not in df.columns]
 if missing_columns:
 raise Exception(f"Missing required columns: {'', '.join(missing_columns)}")

 print("Generating BayesDown syntax with placeholder probabilities...")

 # Build a directed graph of nodes
 G = nx.DiGraph()

 # Add nodes to the graph
 for idx, row in df.iterrows():
 G.add_node(row['Title'], data=row)

```

```

Add edges to the graph based on parent-child relationships - CORRECTLY
for idx, row in df.iterrows():
 child = row['Title']

 # Parse parents and add edges
 parents = row['Parents']
 if isinstance(parents, str):
 # Handle string representation of list
 if parents.startswith('[') and parents.endswith(']'):
 parents = parents.strip('[]')
 if parents: # Check if not empty
 parent_list = [p.strip().strip('\n') for p in parents.split(',')]
 for parent in parent_list:
 if parent in G.nodes():
 # In BayesDown: Parent (cause) -> Child (effect)
 G.add_edge(parent, child)
 elif isinstance(parents, list):
 # Handle actual list
 for parent in parents:
 if parent in G.nodes():
 G.add_edge(parent, child)

Function to safely parse JSON strings
def safe_parse_json(json_str):
 if pd.isna(json_str):
 return {}

 if isinstance(json_str, dict):
 return json_str

 try:
 return json.loads(json_str)
 except:
 return {}

Start building the BayesDown content
bayesdown_content = "" # Initialize as empty

if include_questions_as_comments:
 bayesdown_content = "# BayesDown Representation with Placeholder Probabilities\n\n"
 bayesdown_content += "/* This file contains BayesDown syntax with placeholder probabilities\n\n"
 bayesdown_content += " Replace the placeholders with actual probability values based on"

```

```

 bayesdown_content += " questions in the comments. */\n\n"

Get leaf nodes (nodes with no outgoing edges) - these are effects without children
leaf_nodes = [n for n in G.nodes() if G.out_degree(n) == 0]

Helper function to process a node and its parents recursively
def process_node(node, indent_level=0, processed_nodes=None):
 if processed_nodes is None:
 processed_nodes = set()

 # Create the indentation string
 indent = ' ' * (indent_level * 2)
 prefix = f"{indent}+ " if indent_level > 0 else ""

 # Get node data
 node_data = G.nodes[node]['data']
 title = node_data['Title']
 description = node_data['Description'] if not pd.isna(node_data['Description']) else ""

 # Parse instantiations from the row data
 instantiations = parse_instantiations_safely(node_data['instantiations'])

 # Build the node string
 node_output = ""

 # Add comments with questions if requested
 if include_questions_as_comments:
 # Add positive questions as comments
 if 'Generate_Positive_Instantiation_Questions' in node_data:
 positive_questions = safe_parse_json(node_data['Generate_Positive_Instantiation_Questions'])
 for question in positive_questions.keys():
 node_output += f"{indent}/* {question} */\n"

 # Add negative questions as comments
 if 'Generate_Negative_Instantiation_Questions' in node_data:
 negative_questions = safe_parse_json(node_data['Generate_Negative_Instantiation_Questions'])
 for question in negative_questions.keys():
 node_output += f"{indent}/* {question} */\n"

 # Check if this node was already fully defined elsewhere
 if node in processed_nodes:
 # Just add a reference to the node

```

```

 node_output += f"{prefix}[{title}]\n"
 return node_output

Mark this node as processed
processed_nodes.add(node)

Prepare the metadata JSON
metadata = {
 "instantiations": instantiations
}

Add priors with full questions as keys
priors = {}
if 'Generate_Positive_Instantiation_Questions' in node_data:
 positive_questions = safe_parse_json(node_data['Generate_Positive_Instantiation_Qu
 for question, estimate_key in positive_questions.items():
 if estimate_key == 'prior':
 priors[question] = "%?" # Default placeholder

if 'Generate_Negative_Instantiation_Questions' in node_data:
 negative_questions = safe_parse_json(node_data['Generate_Negative_Instantiation_Qu
 for question, estimate_key in negative_questions.items():
 if estimate_key == 'prior':
 priors[question] = "%?" # Default placeholder

metadata["priors"] = priors

Add posteriors with full questions as keys
parents = list(G.predecessors(node))
if parents:
 posteriors = {}
 if 'Generate_Positive_Instantiation_Questions' in node_data:
 positive_questions = safe_parse_json(node_data['Generate_Positive_Instantiation
 for question, estimate_key in positive_questions.items():
 if estimate_key.startswith('estimate_'):
 posteriors[question] = "?%" # Default placeholder

 if 'Generate_Negative_Instantiation_Questions' in node_data:
 negative_questions = safe_parse_json(node_data['Generate_Negative_Instantiation
 for question, estimate_key in negative_questions.items():
 if estimate_key.startswith('estimate_'):
 posteriors[question] = "?%" # Default placeholder

```

```

 metadata["posteriors"] = posteriors

 # Format the node with metadata
 node_output += f"{prefix}[{title}]: {description} {json.dumps(metadata)}\n"

 # Process parent nodes
 for parent in parents:
 if parent != node: # Avoid self-references
 parent_output = process_node(parent, indent_level + 1, processed_nodes)
 node_output += parent_output

 return node_output

Helper function to parse instantiations safely
def parse_instantiations_safely(instantiations_data):
 if isinstance(instantiations_data, list):
 return instantiations_data if instantiations_data else [f"TRUE", f"FALSE"]

 if isinstance(instantiations_data, str):
 try:
 parsed = json.loads(instantiations_data)
 if isinstance(parsed, list):
 return parsed if parsed else [f"TRUE", f"FALSE"]
 except:
 if instantiations_data.startswith('[') and instantiations_data.endswith(']'):
 items = instantiations_data.strip('[]').split(',')
 result = [item.strip(' "') for item in items if item.strip()]
 return result if result else [f"TRUE", f"FALSE"]

 return [f"TRUE", f"FALSE"] # Default

Process each leaf node and its ancestors
for leaf in leaf_nodes:
 bayesdown_content += process_node(leaf)

Save the BayesDown content
with open(output_md_path, 'w') as f:
 f.write(bayesdown_content)

print(f"BayesDown Questions saved to {output_md_path}")
return bayesdown_content

```

```

Explicitly set the value of include_questions_as_comments
include_questions_as_comments=False # or False, depending on your needs

Get the markdown content
bayesdown_questions = extract_bayesdown_questions_fixed(
 "ArgDown_WithQuestions.csv",
 "BayesDownQuestions.md", include_questions_as_comments=include_questions_as_comments
)

Determine the output file path based on include_questions_as_comments
if include_questions_as_comments: # Assuming include_questions_as_comments is defined somewhere
 output_file_path = "FULL_BayesDownQuestions.md"
else:
 output_file_path = "BayesDownQuestions.md"

Save the markdown content to the appropriate file
with open(output_file_path, 'w') as f:
 f.write(md_content)

print(f"Markdown content saved to {output_file_path}")

```

Loading CSV from ArgDown\_WithQuestions.csv...  
 Successfully loaded CSV with 23 rows.  
 Generating BayesDown syntax with placeholder probabilities...  
 BayesDown Questions saved to BayesDownQuestions.md  
 Markdown content saved to BayesDownQuestions.md

```

Generate BayesDown format
bayesdown_questions = extract_bayesdown_questions_fixed(
 "ArgDown_WithQuestions.csv",
 "FULL_BayesDownQuestions.md",
 include_questions_as_comments=True
)

Display a preview of the format
print("\nBayesDown Format Preview:")
print(bayesdown_questions[:50000] + "...\\n")

```

Loading CSV from ArgDown\_WithQuestions.csv...  
 Successfully loaded CSV with 23 rows.  
 Generating BayesDown syntax with placeholder probabilities...



BayesDown Questions saved to FULL\_BayesDownQuestions.md

BayesDown Format Preview:

# BayesDown Representation with Placeholder Probabilities

```
/* This file contains BayesDown syntax with placeholder probabilities.
 Replace the placeholders with actual probability values based on the
 questions in the comments. */

/* What is the probability for Existential_Catastrophe=existential_catastrophe_TRUE? */
/* What is the probability for Existential_Catastrophe=existential_catastrophe_FALSE? */
[Existential_Catastrophe]: The destruction of humanity's long-term potential due to AI systems
/* What is the probability for Human_Disempowerment=human_disempowerment_TRUE? */
/* What is the probability for Human_Disempowerment=human_disempowerment_TRUE if Scale_Of_Power_Seeking=scale_of_power_seeking_TRUE? */
/* What is the probability for Human_Disempowerment=human_disempowerment_TRUE if Scale_Of_Power_Seeking=scale_of_power_seeking_FALSE? */
/* What is the probability for Human_Disempowerment=human_disempowerment_FALSE? */
/* What is the probability for Human_Disempowerment=human_disempowerment_FALSE if Scale_Of_Power_Seeking=scale_of_power_seeking_TRUE? */
/* What is the probability for Human_Disempowerment=human_disempowerment_FALSE if Scale_Of_Power_Seeking=scale_of_power_seeking_FALSE? */
[Human_Disempowerment]: Permanent and collective disempowerment of humanity relative to AI systems
/* What is the probability for Scale_Of_Power_Seeking=scale_of_power_seeking_TRUE? */
/* What is the probability for Scale_Of_Power_Seeking=scale_of_power_seeking_TRUE if Misaligned_Power_Seeking=misaligned_power_seeking_TRUE? */
/* What is the probability for Scale_Of_Power_Seeking=scale_of_power_seeking_TRUE if Misaligned_Power_Seeking=misaligned_power_seeking_FALSE? */
/* What is the probability for Scale_Of_Power_Seeking=scale_of_power_seeking_FALSE? */
/* What is the probability for Scale_Of_Power_Seeking=scale_of_power_seeking_FALSE if Misaligned_Power_Seeking=misaligned_power_seeking_TRUE? */
/* What is the probability for Scale_Of_Power_Seeking=scale_of_power_seeking_FALSE if Misaligned_Power_Seeking=misaligned_power_seeking_FALSE? */
+ [Scale_Of_Power_Seeking]: Power-seeking by AI systems scaling to the point of permanently disabling humanity
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_TRUE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_TRUE if Existential_Catastrophe=existential_catastrophe_TRUE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_TRUE if Existential_Catastrophe=existential_catastrophe_FALSE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_TRUE if Human_Disempowerment=human_disempowerment_TRUE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_TRUE if Human_Disempowerment=human_disempowerment_FALSE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_TRUE if Scale_Of_Power_Seeking=scale_of_power_seeking_TRUE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_TRUE if Scale_Of_Power_Seeking=scale_of_power_seeking_FALSE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if Existential_Catastrophe=existential_catastrophe_TRUE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if Existential_Catastrophe=existential_catastrophe_FALSE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if Human_Disempowerment=human_disempowerment_TRUE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if Human_Disempowerment=human_disempowerment_FALSE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if Scale_Of_Power_Seeking=scale_of_power_seeking_TRUE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if Scale_Of_Power_Seeking=scale_of_power_seeking_FALSE? */
```

```

/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if
+ [Misaligned_Power_Seeking]: Deployed AI systems seeking power in unintended and high-in
 /* What is the probability for APS_Systems=aps_systems_TRUE? */
 /* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
 /* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
 /* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
 /* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
 /* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
 /* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
 /* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
 /* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
 /* What is the probability for APS_Systems=aps_systems_FALSE? */
 /* What is the probability for APS_Systems=aps_systems_FALSE if Advanced_AI_Capability=
 /* What is the probability for APS_Systems=aps_systems_FALSE if Advanced_AI_Capability=
 /* What is the probability for APS_Systems=aps_systems_FALSE if Advanced_AI_Capability=
 /* What is the probability for APS_Systems=aps_systems_FALSE if Advanced_AI_Capability=
 /* What is the probability for APS_Systems=aps_systems_FALSE if Advanced_AI_Capability=
 /* What is the probability for APS_Systems=aps_systems_FALSE if Advanced_AI_Capability=
 /* What is the probability for APS_Systems=aps_systems_FALSE if Advanced_AI_Capability=
+ [APS_Systems]: AI systems with advanced capabilities, agentic planning, and strategic
 /* What is the probability for Advanced_AI_Capability=advanced_ai_capability_TRUE? */
 /* What is the probability for Advanced_AI_Capability=advanced_ai_capability_FALSE? */
+ [Advanced_AI_Capability]: AI systems that outperform humans on tasks that grant sig
 /* What is the probability for Agentic_Planning=agentic_planning_TRUE? */
 /* What is the probability for Agentic_Planning=agentic_planning_FALSE? */
+ [Agentic_Planning]: AI systems making and executing plans based on world models to
 /* What is the probability for Strategic_Awareness=strategic_awareness_TRUE? */
 /* What is the probability for Strategic_Awareness=strategic_awareness_FALSE? */
+ [Strategic_Awareness]: AI systems with models accurately representing power dynami
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE? */
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if

```

```

/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE? */
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE if Incentives_To_Build_APS=strong? */
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE if Incentives_To_Build_APS=strong? */
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE if Incentives_To_Build_APS=strong? */
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE if Incentives_To_Build_APS=strong? */
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE if Incentives_To_Build_APS=strong? */
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE if Incentives_To_Build_APS=strong? */
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE if Incentives_To_Build_APS=strong? */
+ [Difficulty_Of_Alignment]: It is harder to build aligned systems than misaligned systems. {"instrumental_convergence": true}
/* What is the probability for Instrumental_Convergence=instrumental_convergence_TRUE? */
/* What is the probability for Instrumental_Convergence=instrumental_convergence_FALSE? */
+ [Instrumental_Convergence]: AI systems with misaligned objectives tend to seek power. {"problems_with_proxies": true}
/* What is the probability for Problems_With_Proxies=problems_with_proxies_TRUE? */
/* What is the probability for Problems_With_Proxies=problems_with_proxies_FALSE? */
+ [Problems_With_Proxies]: Optimizing for proxy objectives breaks correlations with the world. {"problems_with_search": true}
/* What is the probability for Problems_With_Search=problems_with_search_TRUE? */
/* What is the probability for Problems_With_Search=problems_with_search_FALSE? */
+ [Problems_With_Search]: Search processes can yield systems pursuing different objectives. {"deployment_decisions": "DEPLOY"}
/* What is the probability for Deployment_Decisions=deployment_decisions_DEPLOY? */
/* What is the probability for Deployment_Decisions=deployment_decisions_DEPLOY if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_DEPLOY if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_DEPLOY if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_DEPLOY if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_DEPLOY if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_DEPLOY if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_DEPLOY if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_DEPLOY if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_WITHHOLD? */
/* What is the probability for Deployment_Decisions=deployment_decisions_WITHHOLD if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_WITHHOLD if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_WITHHOLD if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_WITHHOLD if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_WITHHOLD if Incentives_To_Build_APS=strong? */
+ [Deployment_Decisions]: Decisions to deploy potentially misaligned AI systems. {"incentives_to_build_aps": "STRONG"}
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_STRONG? */
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_STRONG? */
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_STRONG? */
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_STRONG? */
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_STRONG? */
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_STRONG? */
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_STRONG? */
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_STRONG? */
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_WEAK? */
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_WEAK? */
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_WEAK? */
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_WEAK? */
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_WEAK? */
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_WEAK? */
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_WEAK? */
/* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_WEAK? */
+ [Incentives_To_Build_APS]: Strong incentives to build and deploy APS systems. {"instrumental_convergence": true, "problems_with_proxies": true, "problems_with_search": true, "deployment_decisions": "DEPLOY"}
/* What is the probability for Usefulness_Of_APS=usefulness_of_aps_HIGH? */
/* What is the probability for Usefulness_Of_APS=usefulness_of_aps_LOW? */

```



```

/* This file contains BayesDown syntax with placeholder probabilities.
 Replace the placeholders with actual probability values based on the
 questions in the comments. */

/* What is the probability for Existential_Catastrophe=existential_catastrophe_TRUE? */
/* What is the probability for Existential_Catastrophe=existential_catastrophe_FALSE? */
[Existential_Catastrophe]: The destruction of humanity's long-term potential due to AI systems
/* What is the probability for Human_Disempowerment=human_disempowerment_TRUE? */
/* What is the probability for Human_Disempowerment=human_disempowerment_TRUE if Scale_Of_Power_Seeking=scale_of_power_seeking_TRUE? */
/* What is the probability for Human_Disempowerment=human_disempowerment_TRUE if Scale_Of_Power_Seeking=scale_of_power_seeking_FALSE? */
/* What is the probability for Human_Disempowerment=human_disempowerment_FALSE? */
/* What is the probability for Human_Disempowerment=human_disempowerment_FALSE if Scale_Of_Power_Seeking=scale_of_power_seeking_TRUE? */
/* What is the probability for Human_Disempowerment=human_disempowerment_FALSE if Scale_Of_Power_Seeking=scale_of_power_seeking_FALSE? */
[Human_Disempowerment]: Permanent and collective disempowerment of humanity relative to AI systems
/* What is the probability for Scale_Of_Power_Seeking=scale_of_power_seeking_TRUE? */
/* What is the probability for Scale_Of_Power_Seeking=scale_of_power_seeking_TRUE if Misaligned_Power_Seeking=misaligned_power_seeking_TRUE? */
/* What is the probability for Scale_Of_Power_Seeking=scale_of_power_seeking_TRUE if Misaligned_Power_Seeking=misaligned_power_seeking_FALSE? */
/* What is the probability for Scale_Of_Power_Seeking=scale_of_power_seeking_FALSE? */
/* What is the probability for Scale_Of_Power_Seeking=scale_of_power_seeking_FALSE if Misaligned_Power_Seeking=misaligned_power_seeking_TRUE? */
/* What is the probability for Scale_Of_Power_Seeking=scale_of_power_seeking_FALSE if Misaligned_Power_Seeking=misaligned_power_seeking_FALSE? */
+ [Scale_Of_Power_Seeking]: Power-seeking by AI systems scaling to the point of permanently destroying humanity
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_TRUE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_TRUE if Existential_Catastrophe=existential_catastrophe_TRUE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_TRUE if Existential_Catastrophe=existential_catastrophe_FALSE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_TRUE if Human_Disempowerment=human_disempowerment_TRUE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_TRUE if Human_Disempowerment=human_disempowerment_FALSE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_TRUE if Scale_Of_Power_Seeking=scale_of_power_seeking_TRUE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_TRUE if Scale_Of_Power_Seeking=scale_of_power_seeking_FALSE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if Existential_Catastrophe=existential_catastrophe_TRUE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if Existential_Catastrophe=existential_catastrophe_FALSE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if Human_Disempowerment=human_disempowerment_TRUE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if Human_Disempowerment=human_disempowerment_FALSE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if Scale_Of_Power_Seeking=scale_of_power_seeking_TRUE? */
/* What is the probability for Misaligned_Power_Seeking=misaligned_power_seeking_FALSE if Scale_Of_Power_Seeking=scale_of_power_seeking_FALSE? */

```

```

+ [Misaligned_Power_Seeking]: Deployed AI systems seeking power in unintended and high-in
/* What is the probability for APS_Systems=aps_systems_TRUE? */
/* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
/* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
/* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
/* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
/* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
/* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
/* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
/* What is the probability for APS_Systems=aps_systems_TRUE if Advanced_AI_Capability=
/* What is the probability for APS_Systems=aps_systems_FALSE? */
/* What is the probability for APS_Systems=aps_systems_FALSE if Advanced_AI_Capability=
/* What is the probability for APS_Systems=aps_systems_FALSE if Advanced_AI_Capability=
/* What is the probability for APS_Systems=aps_systems_FALSE if Advanced_AI_Capability=
/* What is the probability for APS_Systems=aps_systems_FALSE if Advanced_AI_Capability=
/* What is the probability for APS_Systems=aps_systems_FALSE if Advanced_AI_Capability=
/* What is the probability for APS_Systems=aps_systems_FALSE if Advanced_AI_Capability=
/* What is the probability for APS_Systems=aps_systems_FALSE if Advanced_AI_Capability=
+ [APS_Systems]: AI systems with advanced capabilities, agentic planning, and strategic
/* What is the probability for Advanced_AI_Capability=advanced_ai_capability_TRUE? */
/* What is the probability for Advanced_AI_Capability=advanced_ai_capability_FALSE? */
+ [Advanced_AI_Capability]: AI systems that outperform humans on tasks that grant sig
/* What is the probability for Agentic_Planning=agentic_planning_TRUE? */
/* What is the probability for Agentic_Planning=agentic_planning_FALSE? */
+ [Agentic_Planning]: AI systems making and executing plans based on world models to
/* What is the probability for Strategic_Awareness=strategic_awareness_TRUE? */
/* What is the probability for Strategic_Awareness=strategic_awareness_FALSE? */
+ [Strategic_Awareness]: AI systems with models accurately representing power dynamic
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE? */
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_TRUE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE? */
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE if
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE if

```

```

/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE if Incentives_To_Build_APS=strong? */
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE if Incentives_To_Build_APS=strong? */
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE if Incentives_To_Build_APS=strong? */
/* What is the probability for Difficulty_Of_Alignment=difficulty_of_alignment_FALSE if Incentives_To_Build_APS=strong? */
+ [Difficulty_Of_Alignment]: It is harder to build aligned systems than misaligned systems. {"instantiated": true}
 /* What is the probability for Instrumental_Convergence=instrumental_convergence_TRUE? */
 /* What is the probability for Instrumental_Convergence=instrumental_convergence_FALSE? */
 + [Instrumental_Convergence]: AI systems with misaligned objectives tend to seek power. {"instantiated": true}
 /* What is the probability for Problems_With_Proxies=problems_with_proxies_TRUE? */
 /* What is the probability for Problems_With_Proxies=problems_with_proxies_FALSE? */
 + [Problems_With_Proxies]: Optimizing for proxy objectives breaks correlations with the real world. {"instantiated": true}
 /* What is the probability for Problems_With_Search=problems_with_search_TRUE? */
 /* What is the probability for Problems_With_Search=problems_with_search_FALSE? */
 + [Problems_With_Search]: Search processes can yield systems pursuing different objectives. {"instantiated": true}
/* What is the probability for Deployment_Decisions=deployment_decisions_DEPLOY? */
/* What is the probability for Deployment_Decisions=deployment_decisions_DEPLOY if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_DEPLOY if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_DEPLOY if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_DEPLOY if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_WITHHOLD? */
/* What is the probability for Deployment_Decisions=deployment_decisions_WITHHOLD if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_WITHHOLD if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_WITHHOLD if Incentives_To_Build_APS=strong? */
/* What is the probability for Deployment_Decisions=deployment_decisions_WITHHOLD if Incentives_To_Build_APS=strong? */
+ [Deployment_Decisions]: Decisions to deploy potentially misaligned AI systems. {"instantiated": true}
 /* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_STRONG? */
 /* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_STRONG? */
 /* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_STRONG? */
 /* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_STRONG? */
 /* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_STRONG? */
 /* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_WEAK? */
 /* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_WEAK? */
 /* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_WEAK? */
 /* What is the probability for Incentives_To_Build_APS=incentives_to_build_aps_WEAK? */
 + [Incentives_To_Build_APS]: Strong incentives to build and deploy APS systems. {"instantiated": true}
 /* What is the probability for Usefulness_Of_APS=usefulness_of_aps_HIGH? */
 /* What is the probability for Usefulness_Of_APS=usefulness_of_aps_LOW? */
 + [Usefulness_Of_APS]: APS systems are very useful for many valuable tasks. {"instantiated": true}
 /* What is the probability for Competitive_Dynamics=competitive_dynamics_STRONG? */
 /* What is the probability for Competitive_Dynamics=competitive_dynamics_WEAK? */
 + [Competitive_Dynamics]: Competitive pressures between AI developers. {"instantiated": true}
 /* What is the probability for Deception_By_AI=deception_by_ai_TRUE? */

```

```

 /* What is the probability for Deception_By_AI=deception_by_ai_FALSE? */
 + [Deception_By_AI]: AI systems deceiving humans about their true objectives. {"instant
/* What is the probability for Corrective_Feedback=corrective_feedback_EFFECTIVE? */
/* What is the probability for Corrective_Feedback=corrective_feedback_EFFECTIVE if Warn
/* What is the probability for Corrective_Feedback=corrective_feedback_EFFECTIVE if Warn
/* What is the probability for Corrective_Feedback=corrective_feedback_EFFECTIVE if Warn
/* What is the probability for Corrective_Feedback=corrective_feedback_EFFECTIVE if Warn
/* What is the probability for Corrective_Feedback=corrective_feedback_INEFFECTIVE? */
/* What is the probability for Corrective_Feedback=corrective_feedback_INEFFECTIVE if Wa
/* What is the probability for Corrective_Feedback=corrective_feedback_INEFFECTIVE if Wa
/* What is the probability for Corrective_Feedback=corrective_feedback_INEFFECTIVE if Wa
+ [Corrective_Feedback]: Human society implementing corrections after observing problems
/* What is the probability for Warning_Shots=warning_shots_OBSERVED? */
/* What is the probability for Warning_Shots=warning_shots_UNOBSERVED? */
+ [Warning_Shots]: Observable failures in weaker systems before catastrophic risks. {"
/* What is the probability for Rapid_Capability_Escalation=rapid_capability_escalation
/* What is the probability for Rapid_Capability_Escalation=rapid_capability_escalation
+ [Rapid_Capability_Escalation]: AI capabilities escalating very rapidly, allowing lit
/* What is the probability for Barriers_To_Understanding=barriers_to_understanding_HIGH? */
/* What is the probability for Barriers_To_Understanding=barriers_to_understanding_LOW? */
[Barriers_To_Understanding]: Difficulty in understanding the internal workings of advanced AI
/* What is the probability for Adversarial_Dynamics=adversarial_dynamics_TRUE? */
/* What is the probability for Adversarial_Dynamics=adversarial_dynamics_FALSE? */
[Adversarial_Dynamics]: Potentially adversarial relationships between humans and power-seeking
/* What is the probability for Stakes_Of_Error=stakes_of_error_HIGH? */
/* What is the probability for Stakes_Of_Error=stakes_of_error_LOW? */
[Stakes_Of_Error]: The escalating impact of mistakes with power-seeking AI systems. {"instant

```

```

Generate BayesDown format
bayesdown_questions = extract_bayesdown_questions_fixed(
 "ArgDown_WithQuestions.csv",
 "BayesDownQuestions.md",
 include_questions_as_comments=False
)

Display a preview of the format
print(

)
print(bayesdown_questions[:50000] + "...\\n")

```



```

Loading CSV from ArgDown_WithQuestions.csv...
Successfully loaded CSV with 23 rows.
Generating BayesDown syntax with placeholder probabilities...
BayesDown Questions saved to BayesDownQuestions.md

```

```

[Existential_Catastrophe]: The destruction of humanity's long-term potential due to AI systems. {"instantiation": 1}
[Human_Disempowerment]: Permanent and collective disempowerment of humanity relative to AI systems. {"instantiation": 1}
+ [Scale_Of_Power_Seeking]: Power-seeking by AI systems scaling to the point of permanent domination. {"instantiation": 1}
+ [Misaligned_Power_Seeking]: Deployed AI systems seeking power in unintended and high-risk ways. {"instantiation": 1}
+ [APS_Systems]: AI systems with advanced capabilities, agentic planning, and strategic behavior. {"instantiation": 1}
+ [Advanced_AI_Capability]: AI systems that outperform humans on tasks that grant significant power. {"instantiation": 1}
+ [Agentic_Planning]: AI systems making and executing plans based on world models to achieve long-term goals. {"instantiation": 1}
+ [Strategic_Awareness]: AI systems with models accurately representing power dynamics and human behavior. {"instantiation": 1}
+ [Difficulty_Of_Alignment]: It is harder to build aligned systems than misaligned systems. {"instantiation": 1}
+ [Instrumental_Convergence]: AI systems with misaligned objectives tend to seek power for instrumental reasons. {"instantiation": 1}
+ [Problems_With_Proxies]: Optimizing for proxy objectives breaks correlations with human values. {"instantiation": 1}
+ [Problems_With_Search]: Search processes can yield systems pursuing different objectives than intended. {"instantiation": 1}
+ [Deployment_Decisions]: Decisions to deploy potentially misaligned AI systems. {"instantiation": 1}
+ [Incentives_To_Build_APS]: Strong incentives to build and deploy APS systems. {"instantiation": 1}
+ [Usefulness_Of_APS]: APS systems are very useful for many valuable tasks. {"instantiation": 1}
+ [Competitive_Dynamics]: Competitive pressures between AI developers. {"instantiation": 1}
+ [Deception_By_AI]: AI systems deceiving humans about their true objectives. {"instantiation": 1}
+ [Corrective_Feedback]: Human society implementing corrections after observing problems. {"instantiation": 1}
+ [Warning_Shots]: Observable failures in weaker systems before catastrophic risks. {"instantiation": 1}
+ [Rapid_Capability_Escalation]: AI capabilities escalating very rapidly, allowing little time for correction. {"instantiation": 1}
[Barriers_To_Understanding]: Difficulty in understanding the internal workings of advanced AI systems. {"instantiation": 1}
[Adversarial_Dynamics]: Potentially adversarial relationships between humans and power-seeking AI systems. {"instantiation": 1}
[Stakes_Of_Error]: The escalating impact of mistakes with power-seeking AI systems. {"instantiation": 1}
...

```

## 2.3 Generate BayesDown Probability Extraction Prompt

Generate 2nd Extraction Prompt for Probabilities based on the questions generated from the ‘ArgDown.csv’ extraction

### 2.3.1 BayesDown Format Specification

BayesDown extends ArgDown with probability data in a structured JSON format to represent Bayesian networks. This intermediate representation bridges the gap between natural language arguments and formal probabilistic models, preserving both narrative structure and quantitative relationships.

## Core Structure

A BayesDown representation consists of:

1. **Nodes:** Variables or statements in brackets [Node\_Name] with descriptive text
2. **Relationships:** Hierarchical structure with indentation and + symbols
3. **Metadata:** JSON objects containing probability information:

```
{
 "instantiations": ["state_TRUE", "state_FALSE"], // Possible states of variable
 "priors": {
 "p(state_TRUE)": "0.7", // Unconditional probability of state_TRUE
 "p(state_FALSE)": "0.3" // Unconditional probability of state_FALSE
 },
 "posteriors": {
 "p(state_TRUE|condition1_TRUE,condition2_FALSE)": "0.9", // Conditional on parent states
 "p(state_TRUE|condition1_FALSE,condition2_TRUE)": "0.4" // Different parent configurations
 }
}

Rain-Sprinkler-Lawn Example
[Grass_Wet]: Concentrated moisture on grass. {"instantiations": ["grass_wet_TRUE", "grass_wet_FALSE"],
"priors": {"p(grass_wet_TRUE)": "0.322", "p(grass_wet_FALSE)": "0.678"},
"posteriors": {"p(grass_wet_TRUE|sprinkler_TRUE,rain_TRUE)": "0.99",
"p(grass_wet_TRUE|sprinkler_TRUE,rain_FALSE)": "0.9",
"p(grass_wet_TRUE|sprinkler_FALSE,rain_TRUE)": "0.8",
"p(grass_wet_TRUE|sprinkler_FALSE,rain_FALSE)": "0.0"}}
+ [Rain]: Water falling from the sky. {"instantiations": ["rain_TRUE", "rain_FALSE"],
"priors": {"p(rain_TRUE)": "0.2", "p(rain_FALSE)": "0.8"}}
+ [Sprinkler]: Artificial watering system. {"instantiations": ["sprinkler_TRUE", "sprinkler_FALSE"],
"priors": {"p(sprinkler_TRUE)": "0.44838", "p(sprinkler_FALSE)": "0.55162"},
"posteriors": {"p(sprinkler_TRUE|rain_TRUE)": "0.01", "p(sprinkler_TRUE|rain_FALSE)": "0.4"}
+ [Rain]
```

In this example:

- + Grass\_Wet is the effect/outcome node
- + Rain and Sprinkler are parent nodes (causes)
- + Rain also influences Sprinkler (people tend not to use sprinklers when it's raining)

Role in AMTAIR

BayesDown serves as the critical intermediate representation in the AMTAIR extraction pipeline

For full syntax details, see the BayesDownSyntax.md file in the repository.

### 2.3.2 Probability Extraction Process

The probability extraction pipeline follows these steps:

- Identify variables and their possible states
- Extract prior probability statements
- Identify conditional relationships
- Extract conditional probability statements
- Format the data in BayesDown syntax

### 2.3.3 Implementation Steps

To extract probabilities and create BayesDown format:

- Run the `extract_probabilities` function on ArgDown text
- Process the results into a structured format
- Validate the probability distributions (ensure they sum to 1)
- Generate the enhanced BayesDown representation

### 2.3.4 Validation and Quality Control

The probability extraction process includes validation steps:

- Ensuring coherent probability distributions
- Checking for logical consistency in conditional relationships
- Verifying that all required probability statements are present
- Handling missing data with appropriate default values

## 2.4 Prepare 2nd API call

## 2.5 Make BayesDown Probability Extraction API Call

## 2.6 Save BayesDown with Probability Estimates (.csv)

## 2.7 Review & Verify BayesDown Probability Estimates

## 2.7.2 Check the Graph Structure with the ArgDown Sandbox Online

Copy and paste the BayesDown formatted ... in the ArgDown Sandbox below to quickly verify the

## 2.8 Extract BayesDown with Probability Estimates as Dataframe

# 3.0 Data Extraction: BayesDown (.md) to Database (.csv)

### # 3. BayesDown to Structured Data: Network Construction

#### ## Extraction Pipeline Overview

This section implements the core extraction pipeline described in the AMTAIR project document.

1. **Input**: Text in BayesDown format (see Section 2.3.1)
2. **Parsing**: Extract nodes, relationships, and probability information
3. **Structuring**: Organize into a DataFrame with formal relationships
4. **Enhancement**: Add derived properties and network metrics
5. **Output**: Structured data ready for Bayesian network construction

#### ### Theoretical Foundation

This implementation follows the extraction algorithm outlined in the AMTAIR project description.

1. Get nodes: All premises and conclusions from the argument structure
2. Get edges: Parent-child relationships between nodes
3. Extract probability distributions: Prior and conditional probabilities
4. Calculate derived metrics: Network statistics and node classifications

The resulting structured data maintains the complete information needed to reconstruct the BayesDown input.

#### ### Role in Thesis Research

This extraction pipeline represents a key contribution of the Master's thesis, demonstrating the feasibility of automated argument analysis.

The rain-sprinkler-lawn example serves as a simple but complete test case, demonstrating even the most complex dependencies.

#### ### 3.1 ExtractBayesDown-Data\_v1

Build data frame with extractable information from BayesDown

```
::: {.cell quarto-private-1='{ "key": "colab", "value": { "base_uri": "https://localhost:8080/", "height": 100 } }'}
``` {.python .cell-code}
```

```
# read sprinkler example -- Occam Colab Online
```

```
file_path_ex_rain = "https://raw.githubusercontent.com/SingularitySmith/AMTAIR_Prototype/main/data/ex_rain.json"
```

```
# Use requests.get to fetch content from URL
```

```
response = requests.get(file_path_ex_rain)
```

```
response.raise_for_status() # Raise HTTPError for bad responses (4xx or 5xx)
```

```
# Read content from the response
```

```
md_content_ex_rain = response.text
```

```
md_content_ex_rain
```

```
'[Existential_Catastrophe]: The destruction of humanity\'s long-term potential due to AI sys
```

3.1.2 Test BayesDown Extraction

```
display(Markdown(md_content_ex_rain)) # view BayesDown file formatted as Markdown
```

[Existential_Catastrophe]: The destruction of humanity’s long-term potential due to AI systems we’ve lost control over. {“instantiations”: [“existential_catastrophe_TRUE”, “existential_catastrophe_FALSE”], “priors”: {“p(existential_catastrophe_TRUE)”: “0.05”, “p(existential_catastrophe_FALSE)”: “0.95”}, “posteriors”: {“p(existential_catastrophe_TRUE|human_disempowerment_TRUE)”: “0.95”, “p(existential_catastrophe_TRUE|human_disempowerment_FALSE)”: “0.0”, “p(existential_catastrophe_FALSE|human_disempowerment_TRUE)”: “0.05”, “p(existential_catastrophe_FALSE|human_disempowerment_FALSE)”: “1.0”}} - [Human_Disempowerment]: Permanent and collective disempowerment of humanity relative to AI systems. {“instantiations”: [“human_disempowerment_TRUE”, “human_disempowerment_FALSE”], “priors”: {“p(human_disempowerment_TRUE)”: “0.208”, “p(human_disempowerment_FALSE)”: “0.792”}, “posteriors”: {“p(human_disempowerment_TRUE|scale_of_power_seeking_TRUE)”: “1.0”, “p(human_disempowerment_TRUE|scale_of_power_seeking_FALSE)”: “0.0”, “p(human_disempowerment_FALSE|scale_of_power_seeking_TRUE)”: “0.0”, “p(human_disempowerment_FALSE|scale_of_power_seeking_FALSE)”: “1.0”}} - [Scale_Of_Power_Seeking]: Power-seeking by AI systems scaling to the point of permanently disempowering all of humanity. {“instantiations”: [“scale_of_power_seeking_TRUE”, “scale_of_power_seeking_FALSE”], “priors”: {“p(scale_of_power_seeking_TRUE)”: “0.208”, “p(scale_of_power_seeking_FALSE)”: “0.792”}, “posteriors”: {“p(scale_of_power_seeking_TRUE|corrective_feedback_EFFECTIVE)”: “0.25”, “p(scale_of_power_seeking_TRUE|corrective_feedback_INEFFECTIVE)”: “0.60”, “p(scale_of_power_seeking_FALSE|corrective_feedback_EFFECTIVE)”: “0.0”, “p(scale_of_power_seeking_FALSE|corrective_feedback_INEFFECTIVE)”: “0.0”, “p(scale_of_power_seeking_TRUE|misaligned_power_seeking_TRUE)”: “0.75”, “p(scale_of_power_seeking_TRUE|misaligned_power_seeking_FALSE)”: “0.40”, “p(scale_of_power_seeking_FALSE|misaligned_power_seeking_TRUE)”: “1.0”, “p(scale_of_power_seeking_FALSE|misaligned_power_seeking_FALSE)”: “1.0”}} - [Misaligned_Power_Seeking]: Deployed AI systems seeking power in unintended and high-impact ways due to problems with their objectives. {“instantiations”: [“misaligned_power_seeking_TRUE”, “misaligned_power_seeking_FALSE”], “priors”: {“p(misaligned_power_seeking_TRUE)”:

“0.338”, “p(misaligned_power_seeking_FALSE)” : “0.662”}, “posteriors”: {“p(misaligned_power_seeking_TRUE|difficulty_of_alignment_TRUE, deployment_decisions_DEPLOY)” : “0.90”, “p(misaligned_power_seeking_TRUE|difficulty_of_alignment_TRUE, deployment_decisions_WITHHOLD)” : “0.10”, “p(misaligned_power_seeking_FALSE|difficulty_of_alignment_FALSE, deployment_decisions_DEPLOY)” : “0.25”, “p(misaligned_power_seeking_FALSE|difficulty_of_alignment_FALSE, deployment_decisions_WITHHOLD)” : “0.05”, “p(misaligned_power_seeking_TRUE|difficulty_of_alignment_TRUE, deployment_decisions_DEPLOY)” : “0.0”, “p(misaligned_power_seeking_TRUE|difficulty_of_alignment_TRUE, deployment_decisions_WITHHOLD)” : “0.0”, “p(misaligned_power_seeking_FALSE|difficulty_of_alignment_FALSE, deployment_decisions_DEPLOY)” : “0.0”, “p(misaligned_power_seeking_FALSE|difficulty_of_alignment_FALSE, deployment_decisions_WITHHOLD)” : “0.0”, “p(misaligned_power_seeking_TRUE|difficulty_of_alignment_TRUE, deployment_decisions_DEPLOY)” : “0.10”, “p(misaligned_power_seeking_FALSE|difficulty_of_alignment_TRUE, deployment_decisions_WITHHOLD)” : “0.90”, “p(misaligned_power_seeking_FALSE|difficulty_of_alignment_FALSE, deployment_decisions_DEPLOY)” : “0.75”, “p(misaligned_power_seeking_FALSE|difficulty_of_alignment_FALSE, deployment_decisions_WITHHOLD)” : “0.95”, “p(misaligned_power_seeking_TRUE|difficulty_of_alignment_TRUE, deployment_decisions_DEPLOY)” : “1.0”, “p(misaligned_power_seeking_FALSE|difficulty_of_alignment_TRUE, deployment_decisions_WITHHOLD)” : “1.0”, “p(misaligned_power_seeking_FALSE|difficulty_of_alignment_FALSE, deployment_decisions_DEPLOY)” : “1.0”, “p(misaligned_power_seeking_FALSE|difficulty_of_alignment_FALSE, deployment_decisions_WITHHOLD)” : “1.0”}} - [APS_Systems]:

AI systems with advanced capabilities, agentic planning, and strategic awareness. {“instantiations”: [“aps_systems_TRUE”, “aps_systems_FALSE”], “priors”: {“p(aps_systems_TRUE)” : “0.65”, “p(aps_systems_FALSE)” : “0.35”}, “posteriors”: {“p(aps_systems_TRUE|advanced_ai_capability_TRUE, agentic_planning_TRUE, strategic_awareness_TRUE)” : “1.0”, “p(aps_systems_TRUE|advanced_ai_capability_TRUE, agentic_planning_TRUE, strategic_awareness_FALSE)” : “0.0”, “p(aps_systems_TRUE|advanced_ai_capability_FALSE, agentic_planning_TRUE, strategic_awareness_TRUE)” : “0.0”, “p(aps_systems_TRUE|advanced_ai_capability_FALSE, agentic_planning_FALSE, strategic_awareness_TRUE)” : “0.0”, “p(aps_systems_TRUE|advanced_ai_capability_FALSE, agentic_planning_FALSE, strategic_awareness_FALSE)” : “0.0”, “p(aps_systems_FALSE|advanced_ai_capability_TRUE, agentic_planning_TRUE, strategic_awareness_TRUE)” : “0.0”, “p(aps_systems_FALSE|advanced_ai_capability_TRUE, agentic_planning_TRUE, strategic_awareness_FALSE)” : “0.0”, “p(aps_systems_FALSE|advanced_ai_capability_FALSE, agentic_planning_TRUE, strategic_awareness_TRUE)” : “0.0”, “p(aps_systems_FALSE|advanced_ai_capability_FALSE, agentic_planning_FALSE, strategic_awareness_TRUE)” : “0.0”, “p(aps_systems_FALSE|advanced_ai_capability_FALSE, agentic_planning_FALSE, strategic_awareness_FALSE)” : “0.0”, “p(aps_systems_FALSE|advanced_ai_capability_TRUE, agentic_planning_FALSE, strategic_awareness_TRUE)” : “1.0”, “p(aps_systems_FALSE|advanced_ai_capability_FALSE, agentic_planning_FALSE, strategic_awareness_TRUE)” : “1.0”, “p(aps_systems_FALSE|advanced_ai_capability_FALSE, agentic_planning_FALSE, strategic_awareness_FALSE)” : “1.0”, “p(aps_systems_FALSE|advanced_ai_capability_TRUE, agentic_planning_FALSE, strategic_awareness_TRUE)” : “1.0”, “p(aps_systems_FALSE|advanced_ai_capability_FALSE, agentic_planning_FALSE, strategic_awareness_FALSE)” : “1.0”}} - [Advanced_AI_Capability]:

AI systems that outperform humans on tasks that grant significant power in the world. {“instantiations”: [“advanced_ai_capability_TRUE”, “advanced_ai_capability_FALSE”], “priors”: {“p(advanced_ai_capability_TRUE)” : “0.80”, “p(advanced_ai_capability_FALSE)” : “0.20”}} - [Agentic_Planning]: AI systems making and executing plans based on world models to achieve objectives. {“instantiations”: [“agentic_planning_TRUE”, “agentic_planning_FALSE”], “priors”: {“p(agentic_planning_TRUE)” : “0.85”, “p(agentic_planning_FALSE)” : “0.15”}} - [Strategic_Awareness]: AI systems with models accurately representing

power dynamics with humans. {"instantiations": [{"strategic_awareness_TRUE", "strategic_awareness_FALSE"}], "priors": {"p(strategic_awareness_TRUE)": "0.75", "p(strategic_awareness_FALSE)": "0.25"}} - [Difficulty_Of_Alignment]: It is harder to build aligned systems than misaligned systems that are attractive to deploy. {"instantiations": [{"difficulty_of_alignment_TRUE", "difficulty_of_alignment_FALSE"}], "priors": {"p(difficulty_of_alignment_TRUE)": "0.40", "p(difficulty_of_alignment_FALSE)": "0.60"}}, "posteriors": {"p(difficulty_of_alignment_TRUE|instrumental_convergence_TRUE, problems_with_proxies_TRUE, problems_with_search_TRUE)": "0.85", "p(difficulty_of_alignment_TRUE|problems_with_proxies_TRUE, problems_with_search_FALSE)": "0.70", "p(difficulty_of_alignment_TRUE|problems_with_proxies_FALSE, problems_with_search_TRUE)": "0.60", "p(difficulty_of_alignment_TRUE|problems_with_proxies_FALSE, problems_with_search_FALSE)": "0.40", "p(difficulty_of_alignment_TRUE|problems_with_proxies_TRUE, problems_with_search_TRUE)": "0.55", "p(difficulty_of_alignment_TRUE|problems_with_proxies_TRUE, problems_with_search_FALSE)": "0.40", "p(difficulty_of_alignment_TRUE|problems_with_proxies_FALSE, problems_with_search_TRUE)": "0.30", "p(difficulty_of_alignment_TRUE|problems_with_proxies_FALSE, problems_with_search_FALSE)": "0.10", "p(difficulty_of_alignment_FALSE|problems_with_proxies_TRUE, problems_with_search_TRUE)": "0.15", "p(difficulty_of_alignment_FALSE|problems_with_proxies_TRUE, problems_with_search_FALSE)": "0.30", "p(difficulty_of_alignment_FALSE|problems_with_proxies_FALSE, problems_with_search_TRUE)": "0.40", "p(difficulty_of_alignment_FALSE|problems_with_proxies_FALSE, problems_with_search_FALSE)": "0.60", "p(difficulty_of_alignment_FALSE|problems_with_proxies_TRUE, problems_with_search_TRUE)": "0.45", "p(difficulty_of_alignment_FALSE|problems_with_proxies_TRUE, problems_with_search_FALSE)": "0.60", "p(difficulty_of_alignment_FALSE|problems_with_proxies_FALSE, problems_with_search_TRUE)": "0.70", "p(difficulty_of_alignment_FALSE|problems_with_proxies_FALSE, problems_with_search_FALSE)": "0.90"}}} - [Instrumental_Convergence]: AI systems with misaligned objectives tend to seek power as an instrumental goal. {"instantiations": [{"instrumental_convergence_TRUE", "instrumental_convergence_FALSE"}], "priors": {"p(instrumental_convergence_TRUE)": "0.75", "p(instrumental_convergence_FALSE)": "0.25"}} - [Problems_With_Proxies]: Optimizing for proxy objectives breaks correlations with intended goals. {"instantiations": [{"problems_with_proxies_TRUE", "problems_with_proxies_FALSE"}], "priors": {"p(problems_with_proxies_TRUE)": "0.80", "p(problems_with_proxies_FALSE)": "0.20"}} - [Problems_With_Search]: Search processes can yield systems pursuing different objectives than intended. {"instantiations": [{"problems_with_search_TRUE", "problems_with_search_FALSE"}], "priors": {"p(problems_with_search_TRUE)": "0.70", "p(problems_with_search_FALSE)": "0.30"}} - [Deployment_Decisions]: Decisions to deploy potentially misaligned AI systems. {"instantiations": [{"deployment_decisions_DEPLOY", "deployment_decisions_WITHHOLD"}], "priors": {"p(deployment_decisions_DEPLOY)": "0.70", "p(deployment_decisions_WITHHOLD)": "0.30"}}, "posteriors": {"p(deployment_decisions_DEPLOY|deception_by_ai_TRUE)": "0.90", "p(deployment_decisions_DEPLOY|incentives_to_build_aps_STRONG, deception_by_ai_FALSE)": "0.75", "p(deployment_decisions_DEPLOY|incentives_to_build_aps_WEAK, deception_by_ai_TRUE)": "0.60", "p(deployment_decisions_DEPLOY|incentives_to_build_aps_WEAK, deception_by_ai_FALSE)": "0.30", "p(deployment_decisions_WITHHOLD|incentives_to_build_aps_STRONG, deception_by_ai_TRUE)": "0.10", "p(deployment_decisions_WITHHOLD|incentives_to_build_aps_STRONG, deception_by_ai_FALSE)": "0.25", "p(deployment_decisions_WITHHOLD|incentives_to_build_aps_WEAK, deception_by_ai_TRUE)": "0.10", "p(deployment_decisions_WITHHOLD|incentives_to_build_aps_WEAK, deception_by_ai_FALSE)": "0.25"}}

deception_by_ai_TRUE): “0.40”, “p(deployment_decisions_WITHHOLD|incentives_to_build_aps_WEAK, deception_by_ai_FALSE): “0.70”}} - [Incentives_To_Build_APS]: Strong incentives to build and deploy APS systems. {“instantiations”: [“incentives_to_build_aps_STRONG”, “incentives_to_build_aps_WEAK”], “priors”: {“p(incentives_to_build_aps_STRONG)”: “0.80”, “p(incentives_to_build_aps_WEAK)”: “0.20”}, “posteriors”: {“p(incentives_to_build_aps_STRONG|competitive_dynamics_STRONG)”: “0.95”, “p(incentives_to_build_aps_STRONG|usefulness_of_aps_HIGH, competitive_dynamics_WEAK)”: “0.80”, “p(incentives_to_build_aps_STRONG|usefulness_of_aps_LOW, competitive_dynamics_STRONG)”: “0.70”, “p(incentives_to_build_aps_STRONG|usefulness_of_aps_LOW, competitive_dynamics_WEAK)”: “0.30”, “p(incentives_to_build_aps_WEAK|usefulness_of_aps_HIGH, competitive_dynamics_STRONG)”: “0.05”, “p(incentives_to_build_aps_WEAK|usefulness_of_aps_HIGH, competitive_dynamics_WEAK)”: “0.20”, “p(incentives_to_build_aps_WEAK|usefulness_of_aps_LOW, competitive_dynamics_STRONG)”: “0.30”, “p(incentives_to_build_aps_WEAK|usefulness_of_aps_LOW, competitive_dynamics_WEAK)”: “0.70”}} - [Usefulness_Of_APS]: APS systems are very useful for many valuable tasks. {“instantiations”: [“usefulness_of_aps_HIGH”, “usefulness_of_aps_LOW”], “priors”: {“p(usefulness_of_aps_HIGH)”: “0.85”, “p(usefulness_of_aps_LOW)”: “0.15”}} - [Competitive_Dynamics]: Competitive pressures between AI developers. {“instantiations”: [“competitive_dynamics_STRONG”, “competitive_dynamics_WEAK”], “priors”: {“p(competitive_dynamics_STRONG)”: “0.75”, “p(competitive_dynamics_WEAK)”: “0.25”}} - [Deception_By_AI]: AI systems deceiving humans about their true objectives. {“instantiations”: [“deception_by_ai_TRUE”, “deception_by_ai_FALSE”], “priors”: {“p(deception_by_ai_TRUE)”: “0.50”, “p(deception_by_ai_FALSE)”: “0.50”}} - [Corrective_Feedback]: Human society implementing corrections after observing problems. {“instantiations”: [“corrective_feedback_EFFECTIVE”, “corrective_feedback_INEFFECTIVE”], “priors”: {“p(corrective_feedback_EFFECTIVE)”: “0.60”, “p(corrective_feedback_INEFFECTIVE)”: “0.40”}, “posteriors”: {“p(corrective_feedback_EFFECTIVE|warning_shots_OBSERVED, rapid_capability_escalation_TRUE)”: “0.40”, “p(corrective_feedback_EFFECTIVE|warning_shots_OBSERVED, rapid_capability_escalation_FALSE)”: “0.80”, “p(corrective_feedback_EFFECTIVE|warning_shots_UNOBSERVED, rapid_capability_escalation_TRUE)”: “0.15”, “p(corrective_feedback_EFFECTIVE|warning_shots_UNOBSERVED, rapid_capability_escalation_FALSE)”: “0.50”, “p(corrective_feedback_INEFFECTIVE|warning_shots_OBSERVED, rapid_capability_escalation_TRUE)”: “0.60”, “p(corrective_feedback_INEFFECTIVE|warning_shots_OBSERVED, rapid_capability_escalation_FALSE)”: “0.20”, “p(corrective_feedback_INEFFECTIVE|warning_shots_UNOBSERVED, rapid_capability_escalation_TRUE)”: “0.85”, “p(corrective_feedback_INEFFECTIVE|warning_shots_UNOBSERVED, rapid_capability_escalation_FALSE)”: “0.50”}} - [Warning_Shots]: Observable failures in weaker systems before catastrophic risks. {“instantiations”: [“warning_shots_OBSERVED”, “warning_shots_UNOBSERVED”], “priors”: {“p(warning_shots_OBSERVED)”: “0.70”, “p(warning_shots_UNOBSERVED)”: “0.30”}} - [Rapid_Capability_Escalation]: AI capabilities escalating very rapidly, allowing little time for correction. {“instantiations”: [“rapid_capability_escalation_TRUE”, “rapid_capability_escalation_FALSE”], “priors”: {“p(rapid_capability_escalation_TRUE)”: “0.45”, “p(rapid_capability_escalation_FALSE)”: “0.55”}} [Barriers_To_Understanding]: Difficulty in understanding the internal workings of advanced AI systems. {“instantiations”: [“barriers_to_understanding_HIGH”, “barriers_to_understanding_LOW”], “priors”: {“p(barriers_to_understanding_HIGH)”: “0.70”, “p(barriers_to_understanding_LOW)”: “0.30”}, “posteriors”: {“p(barriers_to_understanding_HIGH|misalignment)”: “0.85”, “p(barriers_to_understanding_LOW|misalignment)”: “0.15”}}

“0.85”, “p(barriers_to_understanding_HIGH|misaligned_power_seeking_FALSE)”:
 “0.60”, “p(barriers_to_understanding_LOW|misaligned_power_seeking_TRUE)”:
 “0.15”, “p(barriers_to_understanding_LOW|misaligned_power_seeking_FALSE)”:
 “0.40”}} - [Misaligned_Power_Seeking]: Deployed AI systems seeking power in unintended and high-impact ways due to problems with their objectives. {“instantiations”: [“misaligned_power_seeking_TRUE”, “misaligned_power_seeking_FALSE”], “priors”: {“p(misaligned_power_seeking_TRUE)”:
 “0.338”, “p(misaligned_power_seeking_FALSE)”:
 “0.662”}} [Adversarial_Dynamics]: Potentially adversarial relationships between humans and power-seeking AI. {“instantiations”: [“adversarial_dynamics_TRUE”, “adversarial_dynamics_FALSE”], “priors”: {“p(adversarial_dynamics_TRUE)”:
 “0.60”, “p(adversarial_dynamics_FALSE)”:
 “0.40”}, “posteriors”: {“p(adversarial_dynamics_TRUE|misaligned_power_seeking_TRUE)”:
 “0.95”, “p(adversarial_dynamics_TRUE|misaligned_power_seeking_FALSE)”:
 “0.10”, “p(adversarial_dynamics_FALSE|misaligned_power_seeking_TRUE)”:
 “0.05”, “p(adversarial_dynamics_FALSE|misaligned_power_seeking_FALSE)”:
 “0.90”}} - [Misaligned_Power_Seeking]: Deployed AI systems seeking power in unintended and high-impact ways due to problems with their objectives. {“instantiations”: [“misaligned_power_seeking_TRUE”, “misaligned_power_seeking_FALSE”], “priors”: {“p(misaligned_power_seeking_TRUE)”:
 “0.338”, “p(misaligned_power_seeking_FALSE)”:
 “0.662”}} [Stakes_Of_Error]: The escalating impact of mistakes with power-seeking AI systems. {“instantiations”: [“stakes_of_error_HIGH”, “stakes_of_error_LOW”], “priors”: {“p(stakes_of_error_HIGH)”:
 “0.85”, “p(stakes_of_error_LOW)”:
 “0.15”}, “posteriors”: {“p(stakes_of_error_HIGH|misaligned_power_seeking_TRUE)”:
 “0.95”, “p(stakes_of_error_HIGH|misaligned_power_seeking_FALSE)”:
 “0.50”, “p(stakes_of_error_LOW|misaligned_power_seeking_TRUE)”:
 “0.05”, “p(stakes_of_error_LOW|misaligned_power_seeking_FALSE)”:
 “0.50”}} - [Misaligned_Power_Seeking]: Deployed AI systems seeking power in unintended and high-impact ways due to problems with their objectives. {“instantiations”: [“misaligned_power_seeking_TRUE”, “misaligned_power_seeking_FALSE”], “priors”: {“p(misaligned_power_seeking_TRUE)”:
 “0.338”, “p(misaligned_power_seeking_FALSE)”:
 “0.662”}}

3.1.2.2 Check the Graph Structure with the ArgDown Sandbox Online

Copy and paste the BayesDown formatted ... in the ArgDown Sandbox below to quickly verify that the network renders correctly.

3.3 Extraction

BayesDown Extraction Code already part of ArgDown extraction code, therefore just use same function “parse_markdown_hierarchy(markdown_data)” and ignore the extra argument (“ArgDown”) because it is automatically set to false and will by default extract BayesDown.

```
result_df = parse_markdown_hierarchy_fixed(md_content_ex_rain)
result_df
```

	Title	Description	line	line_number
0	Existential_Catastrophe	The destruction of humanity's long-term potent...	0	[0]
1	Human_Disempowerment	Permanent and collective disempowerment of hum...	1	[1]
2	Scale_Of_Power_Seeking	Power-seeking by AI systems scaling to the poi...	2	[2]
3	Misaligned_Power_Seeking	Deployed AI systems seeking power in unintende...	3	[3, 21, 23, 25]
4	APS_Systems	AI systems with advanced capabilities, agentic...	4	[4]
5	Advanced_AI_Capability	AI systems that outperform humans on tasks tha...	5	[5]
6	Agentic_Planning	AI systems making and executing plans based on...	6	[6]
7	Strategic_Awareness	AI systems with models accurately representing...	7	[7]
8	Difficulty_Of_Alignment	It is harder to build aligned systems than mis...	8	[8]
9	Instrumental_Convergence	AI systems with misaligned objectives tend to ...	9	[9]
10	Problems_With_Proxies	Optimizing for proxy objectives breaks correla...	10	[10]
11	Problems_With_Search	Search processes can yield systems pursuing di...	11	[11]
12	Deployment_Decisions	Decisions to deploy potentially misaligned AI ...	12	[12]
13	Incentives_To_Build_APS	Strong incentives to build and deploy APS syst...	13	[13]
14	Usefulness_Of_APS	APS systems are very useful for many valuable ...	14	[14]
15	Competitive_Dynamics	Competitive pressures between AI developers.	15	[15]
16	Deception_By_AI	AI systems deceiving humans about their true o...	16	[16]
17	Corrective_Feedback	Human society implementing corrections after o...	17	[17]
18	Warning_Shots	Observable failures in weaker systems before c...	18	[18]
19	Rapid_Capability_Escalation	AI capabilities escalating very rapidly, allow...	19	[19]
20	Barriers_To_Understanding	Difficulty in understanding the internal worki...	20	[20]
21	Adversarial_Dynamics	Potentially adversarial relationships between ...	22	[22]
22	Stakes_Of_Error	The escalating impact of mistakes with power-s...	24	[24]

3.3 Data-Post-Processing

Add rows to data frame that can be calculated from the extracted rows

```
# @title 3.3.1 Data Post-Processing Functions ---

"""
BLOCK PURPOSE: Enhances the extracted BayesDown data with calculated metrics and network properties.

This block provides functions to enrich the basic extracted data with additional
calculated columns that are useful for analysis and visualization:

1. Joint probabilities - Calculating P(A,B) from conditional and prior probabilities
2. Network metrics - Centrality measures that indicate importance of nodes in the network
3. Markov blanket - Identifying the minimal set of nodes that shield a node from the rest
```

These enhancements provide valuable context for understanding the network structure and the relationships between variables, enabling more advanced analysis and improving visualization.

DEPENDENCIES: networkx for graph calculations

INPUTS: DataFrame with basic extracted BayesDown data

OUTPUTS: Enhanced DataFrame with additional calculated columns

"""

```
def enhance_extracted_data(df):
```

```
    """
```

```
    Enhance the extracted data with calculated columns
```

```
    Args:
```

```
        df: DataFrame with extracted BayesDown data
```

```
    Returns:
```

```
        Enhanced DataFrame with additional columns
```

```
    """
```

```
    # Create a copy to avoid modifying the original
```

```
    enhanced_df = df.copy()
```

```
    # 1. Calculate joint probabilities -  $P(A,B) = P(A|B) * P(B)$ 
```

```
    enhanced_df['joint_probabilities'] = None
```

```
    for idx, row in enhanced_df.iterrows():
```

```
        title = row['Title']
```

```
        priors = row['priors'] if isinstance(row['priors'], dict) else {}
```

```
        posteriors = row['posteriors'] if isinstance(row['posteriors'], dict) else {}
```

```
        parents = row['Parents'] if isinstance(row['Parents'], list) else []
```

```
    # Skip if no parents or no priors
```

```
    if not parents or not priors:
```

```
        continue
```

```
    # Initialize joint probabilities dictionary
```

```
    joint_probs = {}
```

```
    # Get instantiations
```

```
    instantiations = row['instantiations']
```

```
    if not isinstance(instantiations, list) or not instantiations:
```

```
        continue
```

```

# For each parent and child instantiation combination, calculate joint probability
for inst in instantiations:
    # Get this instantiation's prior probability
    inst_prior_key = f"p({inst})"
    if inst_prior_key not in priors:
        continue

    try:
        inst_prior = float(priors[inst_prior_key])
    except (ValueError, TypeError):
        continue

    # For each parent
    for parent in parents:
        parent_row = enhanced_df[enhanced_df['Title'] == parent]
        if parent_row.empty:
            continue

        parent_insts = parent_row.iloc[0]['instantiations']
        if not isinstance(parent_insts, list) or not parent_insts:
            continue

        for parent_inst in parent_insts:
            # Get conditional probability
            cond_key = f"p({inst}|{parent}={parent_inst})"
            if cond_key in posteriors:
                try:
                    cond_prob = float(posteriors[cond_key])

                    # Get parent's prior
                    parent_priors = parent_row.iloc[0]['priors']
                    if not isinstance(parent_priors, dict):
                        continue

                    parent_prior_key = f"p({parent_inst})"
                    if parent_prior_key not in parent_priors:
                        continue

                    try:
                        parent_prior = float(parent_priors[parent_prior_key])

                        # Calculate joint probability: P(A,B) = P(A|B) * P(B)

```

```

        joint_prob = cond_prob * parent_prior
        joint_key = f"p({inst},{parent}={parent_inst})"
        joint_probs[joint_key] = str(round(joint_prob, 4))
    except (ValueError, TypeError):
        joint_prob = cond_prob * parent_prior
        joint_key = f"p({inst},{parent}={parent_inst})"
        joint_probs[joint_key] = str(round(joint_prob, 4))
    except (ValueError, TypeError):
        continue
except (ValueError, TypeError):
    continue

# Store joint probabilities in dataframe
enhanced_df.at[idx, 'joint_probabilities'] = joint_probs

# 2. Calculate network metrics
# Create a directed graph
import networkx as nx
G = nx.DiGraph()

# Add nodes
for idx, row in enhanced_df.iterrows():
    G.add_node(row['Title'])

# Add edges
for idx, row in enhanced_df.iterrows():
    child = row['Title']
    parents = row['Parents'] if isinstance(row['Parents'], list) else []

    for parent in parents:
        if parent in G.nodes():
            G.add_edge(parent, child)

# Calculate centrality measures
degree centrality = nx.degree_centrality(G) # Overall connectedness
in_degree centrality = nx.in_degree_centrality(G) # How many nodes affect this one
out_degree centrality = nx.out_degree_centrality(G) # How many nodes this one affects

try:
    betweenness centrality = nx.betweenness_centrality(G) # Node's role as a connector
except:
    betweenness centrality = {node: 0 for node in G.nodes()}

```

```

# Add metrics to dataframe
enhanced_df['degree centrality'] = None
enhanced_df['in_degree centrality'] = None
enhanced_df['out_degree centrality'] = None
enhanced_df['betweenness centrality'] = None

for idx, row in enhanced_df.iterrows():
    title = row['Title']
    enhanced_df.at[idx, 'degree centrality'] = degree centrality.get(title, 0)
    enhanced_df.at[idx, 'in_degree centrality'] = in_degree centrality.get(title, 0)
    enhanced_df.at[idx, 'out_degree centrality'] = out_degree centrality.get(title, 0)
    enhanced_df.at[idx, 'betweenness centrality'] = betweenness centrality.get(title, 0)

# 3. Add Markov blanket information (parents, children, and children's parents)
enhanced_df['markov_blanket'] = None

for idx, row in enhanced_df.iterrows():
    title = row['Title']
    parents = row['Parents'] if isinstance(row['Parents'], list) else []
    children = row['Children'] if isinstance(row['Children'], list) else []

    # Get children's parents (excluding this node)
    childrens_parents = []
    for child in children:
        child_row = enhanced_df[enhanced_df['Title'] == child]
        if not child_row.empty:
            child_parents = child_row.iloc[0]['Parents']
            if isinstance(child_parents, list):
                childrens_parents.extend([p for p in child_parents if p != title])

    # Remove duplicates
    childrens_parents = list(set(childrens_parents))

    # Combine to get Markov blanket
    markov_blanket = list(set(parents + children + childrens_parents))
    enhanced_df.at[idx, 'markov_blanket'] = markov_blanket

return enhanced_df

```

```

# @title 3.3 --- Enhance Extracted Data with Network Metrics ---

```

```

"""
BLOCK PURPOSE: Applies the post-processing functions to enhance the extracted data.

This block takes the basic extracted DataFrame from the BayesDown parsing step
and enriches it with calculated metrics that provide deeper insight into the
network structure and relationships. It:

1. Applies the enhancement functions defined previously
2. Displays summary information about key calculated metrics
3. Saves the enhanced data for further analysis and visualization

The enhanced DataFrame provides a richer representation of the Bayesian network,
including measures of node importance and conditional relationships that are
essential for effective analysis and visualization.

DEPENDENCIES: enhance_extracted_data function
INPUTS: DataFrame with basic extracted BayesDown data
OUTPUTS: Enhanced DataFrame with additional calculated columns, saved to CSV
"""

# Enhance the extracted dataframe with calculated columns
enhanced_df = enhance_extracted_data(result_df)

# Display the enhanced dataframe
print("Enhanced DataFrame with additional calculated columns:")
enhanced_df.head()

# Check some calculated metrics
print("\nJoint Probabilities Example:")
example_node = enhanced_df.loc[0, 'Title']
joint_probs = enhanced_df.loc[0, 'joint_probabilities']
print(f"Joint probabilities for {example_node}:")
print(joint_probs)

print("\nNetwork Metrics:")
for idx, row in enhanced_df.iterrows():
    print(f"{row['Title']}:")
    print(f"  Degree Centrality: {row['degree_centrality']:.3f}")
    print(f"  Betweenness Centrality: {row['betweenness_centrality']:.3f}")

# Save the enhanced dataframe
enhanced_df.to_csv('enhanced_extracted_data.csv', index=False)

```

```
print("\nEnhanced data saved to 'enhanced_extracted_data.csv'")
```

Enhanced DataFrame with additional calculated columns:

Joint Probabilities Example:

Joint probabilities for Existential_Catastrophe:

None

Network Metrics:

Existential_Catastrophe:

Degree Centrality: 0.000

Betweenness Centrality: 0.000

Human_Disempowerment:

Degree Centrality: 0.045

Betweenness Centrality: 0.000

Scale_Of_Power_Seeking:

Degree Centrality: 0.136

Betweenness Centrality: 0.037

Misaligned_Power_Seeking:

Degree Centrality: 0.182

Betweenness Centrality: 0.056

APS_Systems:

Degree Centrality: 0.182

Betweenness Centrality: 0.019

Advanced_AI_Capability:

Degree Centrality: 0.045

Betweenness Centrality: 0.000

Agentic_Planning:

Degree Centrality: 0.045

Betweenness Centrality: 0.000

Strategic_Awareness:

Degree Centrality: 0.045

Betweenness Centrality: 0.000

Difficulty_Of_Alignment:

Degree Centrality: 0.182

Betweenness Centrality: 0.019

Instrumental_Convergence:

Degree Centrality: 0.045

Betweenness Centrality: 0.000

Problems_With_Proxies:

Degree Centrality: 0.045

Betweenness Centrality: 0.000

Problems_With_Search:
Degree Centrality: 0.045
Betweenness Centrality: 0.000
Deployment_Decisions:
Degree Centrality: 0.136
Betweenness Centrality: 0.026
Incentives_To_Build_APS:
Degree Centrality: 0.136
Betweenness Centrality: 0.017
Usefulness_Of_APS:
Degree Centrality: 0.045
Betweenness Centrality: 0.000
Competitive_Dynamics:
Degree Centrality: 0.045
Betweenness Centrality: 0.000
Deception_By_AI:
Degree Centrality: 0.045
Betweenness Centrality: 0.000
Corrective_Feedback:
Degree Centrality: 0.136
Betweenness Centrality: 0.009
Warning_Shots:
Degree Centrality: 0.045
Betweenness Centrality: 0.000
Rapid_Capability_Escalation:
Degree Centrality: 0.045
Betweenness Centrality: 0.000
Barriers_To_Understanding:
Degree Centrality: 0.000
Betweenness Centrality: 0.000
Adversarial_Dynamics:
Degree Centrality: 0.000
Betweenness Centrality: 0.000
Stakes_Of_Error:
Degree Centrality: 0.000
Betweenness Centrality: 0.000

Enhanced data saved to 'enhanced_extracted_data.csv'

3.4 Download and save finished data frame as .csv file

```
# @title 3.4 --- Save Extracted Data for Further Processing ---

"""
BLOCK PURPOSE: Saves the extracted data to a CSV file for further processing.

This step is essential for:
1. Persisting the structured representation of the Bayesian network
2. Enabling further analysis in other tools or notebook sections
3. Creating a permanent record of the extraction results
4. Making the data available for the visualization pipeline

The CSV format provides a standardized, tabular representation of the network
that can be easily loaded and processed in subsequent analysis steps.

DEPENDENCIES: pandas DataFrame operations
INPUTS: Extracted DataFrame from the parsing step
OUTPUTS: CSV file containing the structured network data
"""

# Save the extracted data as a CSV file
result_df.to_csv('extracted_data.csv', index=False)

print(" Extracted data saved successfully to 'extracted_data.csv'")
print("Note: If using updated data in future steps, the file must be pushed to the GitHub repository")
```

```
Extracted data saved successfully to 'extracted_data.csv'
Note: If using updated data in future steps, the file must be pushed to the GitHub repository
```

4. Analysis & Inference: Bayesian Network Visualization

Bayesian Network Visualization Approach

This section implements the visualization component of the AMTAIR project, transforming the structured data extracted from BayesDown into an interactive network visualization that makes complex probabilistic relationships accessible to human understanding.

Visualization Philosophy

A key challenge in AI governance is making complex probabilistic relationships understandable to diverse stakeholders. This visualization system addresses this challenge through:

1. **Visual Encoding of Probability:** Node colors reflect probability values (green for high probability, red for low)
2. **Structural Classification:** Border colors indicate node types (blue for root causes, purple for intermediate nodes, magenta for leaf nodes)
3. **Progressive Disclosure:** Basic information in tooltips, detailed probability tables in modal popups
4. **Interactive Exploration:** Draggable nodes, configurable physics, click interactions

Connection to AMTAIR Goals

This visualization approach directly supports the AMTAIR project's goal of improving coordination in AI governance by:

1. Making implicit models explicit through visual representation
2. Providing a common language for discussing probabilistic relationships
3. Enabling non-technical stakeholders to engage with formal models
4. Creating shareable artifacts that facilitate collaboration

Implementation Structure

The visualization system is implemented in four phases:

1. **Network Construction:** Creating a directed graph representation using NetworkX
2. **Node Classification:** Identifying node types based on network position
3. **Visual Enhancement:** Adding color coding, tooltips, and interactive elements
4. **Interactive Features:** Implementing click handling for detailed exploration

The resulting visualization serves as both an analytical tool for experts and a communication tool for broader audiences, bridging the gap between technical and policy domains in AI governance discussions.

Phase 1: Dependencies/Functions

```

# @title 4.0 --- Bayesian Network Visualization Functions ---

"""
BLOCK PURPOSE: Provides functions to create interactive Bayesian network visualizations
from DataFrame representations of ArgDown/BayesDown data.

This block implements the visualization pipeline described in the AMTAIR project, transforming
the structured DataFrame extracted from ArgDown/BayesDown into an interactive network graph
that displays nodes, relationships, and probability information. The visualization leverages
NetworkX for graph representation and PyVis for interactive display.

Key visualization features:
1. Color-coding of nodes based on probability values
2. Border styling to indicate node types (root, intermediate, leaf)
3. Interactive tooltips with probability information
4. Modal popups with detailed conditional probability tables
5. Physics-based layout for intuitive exploration

DEPENDENCIES: networkx, pyvis, HTML display from IPython
INPUTS: DataFrame with node information, relationships, and probabilities
OUTPUTS: Interactive HTML visualization of the Bayesian network
"""

from pyvis.network import Network
import networkx as nx
from IPython.display import HTML
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import io
import base64
import colorsys
import json

def create_bayesian_network_with_probabilities(df):
    """
    Create an interactive Bayesian network visualization with enhanced probability visualization
    and node classification based on network structure.

    Args:
        df (pandas.DataFrame): DataFrame containing node information, relationships, and probabilities

```

```

Returns:
    IPython.display.HTML: Interactive HTML visualization of the Bayesian network
    """
# PHASE 1: Create a directed graph representation
G = nx.DiGraph()

# Add nodes with proper attributes
for idx, row in df.iterrows():
    title = row['Title']
    description = row['Description']

    # Process probability information
    priors = get_priors(row)
    instantiations = get_instantiations(row)

    # Add node with base information
    G.add_node(
        title,
        description=description,
        priors=priors,
        instantiations=instantiations,
        posteriors=get_posteriors(row)
    )

# Add edges based on parent-child relationships
for idx, row in df.iterrows():
    child = row['Title']
    parents = get_parents(row)

    # Add edges from each parent to this child
    for parent in parents:
        if parent in G.nodes():
            G.add_edge(parent, child)

# PHASE 2: Classify nodes based on network structure
classify_nodes(G)

# PHASE 3: Create interactive network visualization
net = Network(notebook=True, directed=True, cdn_resources="in_line", height="600px", width="1000px")

# Configure physics for better layout
net.force_atlas_2based(gravity=-50, spring_length=100, spring_strength=0.02)

```

```

net.show_buttons(filter_=['physics']) # Allow user to adjust physics settings

# Add the graph to the network
net.from_nx(G)

# PHASE 4: Enhance node appearance with probability information
for node in net.nodes:
    node_id = node['id']
    node_data = G.nodes[node_id]

    # Get node type and set border color
    node_type = node_data.get('node_type', 'unknown')
    border_color = get_border_color(node_type)

    # Get probability information
    priors = node_data.get('priors', {})
    true_prob = priors.get('true_prob', 0.5) if priors else 0.5

    # Get proper state names
    instantiations = node_data.get('instantiations', ["TRUE", "FALSE"])
    true_state = instantiations[0] if len(instantiations) > 0 else "TRUE"
    false_state = instantiations[1] if len(instantiations) > 1 else "FALSE"

    # Create background color based on probability
    background_color = get_probability_color(priors)

    # Create tooltip with probability information
    tooltip = create_tooltip(node_id, node_data)

    # Create a simpler node label with probability
    simple_label = f"{node_id}\n{true_prob:.2f}"

    # Store expanded content as a node attribute for use in click handler
    node_data['expanded_content'] = create_expanded_content(node_id, node_data)

    # Set node attributes
    node['title'] = tooltip # Tooltip HTML
    node['label'] = simple_label # Simple text label
    node['shape'] = 'box'
    node['color'] = {
        'background': background_color,
        'border': border_color,

```

```

        'highlight': {
            'background': background_color,
            'border': border_color
        }
    }

# PHASE 5: Setup interactive click handling
# Prepare data for click handler
setup_data = {
    'nodes_data': {node_id: {
        'expanded_content': json.dumps(G.nodes[node_id].get('expanded_content', '')),
        'description': G.nodes[node_id].get('description', ''),
        'priors': G.nodes[node_id].get('priors', {}),
        'posteriors': G.nodes[node_id].get('posteriors', {})
    } for node_id in G.nodes()}
}

# JavaScript code for handling node clicks
click_js = """
// Store node data for click handling
var nodesData = %s;

// Add event listener for node clicks
network.on("click", function(params) {
    if (params.nodes.length > 0) {
        var nodeId = params.nodes[0];
        var nodeInfo = nodesData[nodeId];

        if (nodeInfo) {
            // Create a modal popup for expanded content
            var modal = document.createElement('div');
            modal.style.position = 'fixed';
            modal.style.left = '50%%';
            modal.style.top = '50%%';
            modal.style.transform = 'translate(-50%%, -50%%)';
            modal.style.backgroundColor = 'white';
            modal.style.padding = '20px';
            modal.style.borderRadius = '5px';
            modal.style.boxShadow = '0 0 10px rgba(0,0,0,0.5)';
            modal.style.zIndex = '1000';
            modal.style.maxWidth = '80%%';
            modal.style.maxHeight = '80%%';

```

```

        modal.style.overflow = 'auto';

        // Add expanded content
        modal.innerHTML = nodeInfo.expanded_content || 'No detailed information available';

        // Add close button
        var closeBtn = document.createElement('button');
        closeBtn.innerHTML = 'Close';
        closeBtn.style.marginTop = '10px';
        closeBtn.style.padding = '5px 10px';
        closeBtn.style.cursor = 'pointer';
        closeBtn.onclick = function() {
            document.body.removeChild(modal);
        };
        modal.appendChild(closeBtn);

        // Add modal to body
        document.body.appendChild(modal);
    }
}

});
""" % json.dumps(setup_data['nodes_data'])

# PHASE 6: Save the graph to HTML and inject custom click handling
html_file = "bayesian_network.html"
net.save_graph(html_file)

# Inject custom click handling into HTML
try:
    with open(html_file, "r") as f:
        html_content = f.read()

    # Insert click handling script before the closing body tag
    html_content = html_content.replace('</body>', f'<script>{click_js}</script></body>')

    # Write back the modified HTML
    with open(html_file, "w") as f:
        f.write(html_content)

    return HTML(html_content)
except Exception as e:
    return HTML(f"<p>Error rendering HTML: {str(e)}</p><p>The network visualization has failed to render.</p>")

```


Phase 2: Node Classification and Styling Module

```
# @title 4.1 --- Node Classification and Styling Functions ---

"""
BLOCK PURPOSE: Implements the visual classification and styling of nodes in the Bayesian network

This module handles the identification of node types based on their position in the network
and provides appropriate visual styling for each type. The functions:

1. Classify nodes as parents (causes), children (intermediate effects), or leaves (final effects)
2. Assign appropriate border colors to visually distinguish node types
3. Calculate background colors based on probability values
4. Extract relevant information from DataFrame rows in a robust manner

The visual encoding helps users understand both the structure of the network
and the probability distributions at a glance.

DEPENDENCIES: colorsys for color manipulation
INPUTS: Graph structure and node data
OUTPUTS: Classification and styling information for visualization
"""

def classify_nodes(G):
    """
    Classify nodes as parent, child, or leaf based on network structure

    Args:
        G (networkx.DiGraph): Directed graph representation of the Bayesian network

    Effects:
        Adds 'node_type' attribute to each node in the graph:
        - 'parent': Root node with no parents but has children (causal source)
        - 'child': Node with both parents and children (intermediate)
        - 'leaf': Node with parents but no children (final effect)
        - 'isolated': Node with no connections (rare in Bayesian networks)
    """
    for node in G.nodes():
        predecessors = list(G.predecessors(node)) # Nodes pointing to this one (causes)
        successors = list(G.successors(node))     # Nodes this one points to (effects)

        if not predecessors: # No parents
```

```

        if successors: # Has children
            G.nodes[node]['node_type'] = 'parent' # Root cause
        else: # No children either
            G.nodes[node]['node_type'] = 'isolated' # Disconnected node
    else: # Has parents
        if not successors: # No children
            G.nodes[node]['node_type'] = 'leaf' # Final effect
        else: # Has both parents and children
            G.nodes[node]['node_type'] = 'child' # Intermediate node

def get_border_color(node_type):
    """
    Return border color based on node type

    Args:
        node_type (str): Type of node ('parent', 'child', 'leaf', or 'isolated')

    Returns:
        str: Hex color code for node border
    """
    if node_type == 'parent':
        return '#0000FF' # Blue for root causes
    elif node_type == 'child':
        return '#800080' # Purple for intermediate nodes
    elif node_type == 'leaf':
        return '#FF00FF' # Magenta for final effects
    else:
        return '#000000' # Default black for any other type

def get_probability_color(priors):
    """
    Create background color based on probability (red to green gradient)

    Args:
        priors (dict): Dictionary containing probability information

    Returns:
        str: Hex color code for node background, ranging from red (low probability)
            to green (high probability)
    """
    # Default to neutral color if no probability
    if not priors or 'true_prob' not in priors:

```

```

        return '#F8F8F8' # Light grey

# Get probability value
prob = priors['true_prob']

# Create color gradient from red (0.0) to green (1.0)
hue = 120 * prob # 0 = red, 120 = green (in HSL color space)
saturation = 0.75
lightness = 0.8 # Lighter color for better text visibility

# Convert HSL to RGB
r, g, b = colorsys.hls_to_rgb(hue/360, lightness, saturation)

# Convert to hex format
hex_color = "#{:02x}{:02x}{:02x}".format(int(r*255), int(g*255), int(b*255))

return hex_color

def get_parents(row):
    """
    Extract parent nodes from row data, with safe handling for different data types

    Args:
        row (pandas.Series): Row from DataFrame containing node information

    Returns:
        list: List of parent node names
    """
    if 'Parents' not in row:
        return []

    parents_data = row['Parents']

    # Handle NaN, None, or empty list
    if isinstance(parents_data, float) and pd.isna(parents_data):
        return []

    if parents_data is None:
        return []

    # Handle different data types
    if isinstance(parents_data, list):

```

```

        # Return a list with NaN and empty strings removed
        return [p for p in parents_data if not (isinstance(p, float) and pd.isna(p)) and p != '']

    if isinstance(parents_data, str):
        if not parents_data.strip():
            return []

        # Remove brackets and split by comma, removing empty strings and NaN
        cleaned = parents_data.strip('[]"\''')
        if not cleaned:
            return []

        return [p.strip(' "') for p in cleaned.split(',') if p.strip()]

    # Default: empty list
    return []

def get_instantiations(row):
    """
    Extract instantiations with safe handling for different data types

    Args:
        row (pandas.Series): Row from DataFrame containing node information

    Returns:
        list: List of possible instantiations (states) for the node
    """
    if 'instantiations' not in row:
        return ["TRUE", "FALSE"]

    inst_data = row['instantiations']

    # Handle NaN or None
    if isinstance(inst_data, float) and pd.isna(inst_data):
        return ["TRUE", "FALSE"]

    if inst_data is None:
        return ["TRUE", "FALSE"]

    # Handle different data types
    if isinstance(inst_data, list):
        return inst_data if inst_data else ["TRUE", "FALSE"]

```

```

if isinstance(inst_data, str):
    if not inst_data.strip():
        return ["TRUE", "FALSE"]

    # Remove brackets and split by comma
    cleaned = inst_data.strip('[]"\'')
    if not cleaned:
        return ["TRUE", "FALSE"]

    return [i.strip(' "') for i in cleaned.split(',') if i.strip()]

# Default
return ["TRUE", "FALSE"]

def get_priors(row):
    """
    Extract prior probabilities with safe handling for different data types

    Args:
        row (pandas.Series): Row from DataFrame containing node information

    Returns:
        dict: Dictionary of prior probabilities with 'true_prob' added for convenience
    """
    if 'priors' not in row:
        return {}

    priors_data = row['priors']

    # Handle NaN or None
    if isinstance(priors_data, float) and pd.isna(priors_data):
        return {}

    if priors_data is None:
        return {}

    result = {}

    # Handle dictionary
    if isinstance(priors_data, dict):
        result = priors_data
    # Handle string representation of dictionary

```

```

elif isinstance(priors_data, str):
    if not priors_data.strip() or priors_data == '{}':
        return {}

    try:
        # Try to evaluate as Python literal
        import ast
        result = ast.literal_eval(priors_data)
    except:
        # Simple parsing for items like {'p(TRUE)': '0.2', 'p(FALSE)': '0.8'}
        if '{' in priors_data and '}' in priors_data:
            content = priors_data[priors_data.find('{')+1:priors_data.rfind('}')]
            items = [item.strip() for item in content.split(',')]

            for item in items:
                if ':' in item:
                    key, value = item.split(':', 1)
                    key = key.strip(' \\'')
                    value = value.strip(' \\'')
                    result[key] = value

# Extract main probability for TRUE state
instantiations = get_instantiations(row)
true_state = instantiations[0] if instantiations else "TRUE"
true_key = f"p({true_state})"

if true_key in result:
    try:
        result['true_prob'] = float(result[true_key])
    except:
        pass

return result

def get_posteriors(row):
    """
    Extract posterior probabilities with safe handling for different data types

    Args:
        row (pandas.Series): Row from DataFrame containing node information

    Returns:

```

```

    dict: Dictionary of conditional probabilities
    """
    if 'posteriors' not in row:
        return {}

    posteriors_data = row['posteriors']

    # Handle NaN or None
    if isinstance(posteriors_data, float) and pd.isna(posteriors_data):
        return {}

    if posteriors_data is None:
        return {}

    result = {}

    # Handle dictionary
    if isinstance(posteriors_data, dict):
        result = posteriors_data
    # Handle string representation of dictionary
    elif isinstance(posteriors_data, str):
        if not posteriors_data.strip() or posteriors_data == '{}':
            return {}

        try:
            # Try to evaluate as Python literal
            import ast
            result = ast.literal_eval(posteriors_data)
        except:
            # Simple parsing
            if '{' in posteriors_data and '}' in posteriors_data:
                content = posteriors_data[posteriors_data.find('{')+1:posteriors_data.rfind('}')]
                items = [item.strip() for item in content.split(',')]

                for item in items:
                    if ':' in item:
                        key, value = item.split(':', 1)
                        key = key.strip(' \\'\"')
                        value = value.strip(' \\'\"')
                        result[key] = value

    return result

```

Phase 3: HTML Content Generation Module

```
# @title 4.2 --- HTML Content Generation Functions ---

"""
BLOCK PURPOSE: Creates rich HTML content for the interactive Bayesian network visualization.

This module generates the HTML components that enhance the Bayesian network visualization:
1. Probability bars - Visual representation of probability distributions
2. Node tooltips - Rich information displayed on hover
3. Expanded content - Detailed probability information shown when clicking nodes

These HTML components make the mathematical concepts of Bayesian networks more
intuitive and accessible to users without requiring deep statistical knowledge.
The visual encoding of probabilities (colors, bars) and the progressive disclosure
of information (hover, click) help users build understanding at their own pace.

DEPENDENCIES: HTML generation capabilities
INPUTS: Node data from the Bayesian network
OUTPUTS: HTML content for visualization components
"""

def create_probability_bar(true_prob, false_prob, height="15px", show_values=True, value_prefix="p="):
    """
    Creates a reusable HTML component to visualize probability distribution

    Args:
        true_prob (float): Probability of the true state (0.0-1.0)
        false_prob (float): Probability of the false state (0.0-1.0)
        height (str): CSS height of the bar
        show_values (bool): Whether to display numerical values
        value_prefix (str): Prefix to add before values (e.g., "p=")

    Returns:
        str: HTML for a horizontal bar showing probabilities
    """
    # Prepare display labels if showing values
    true_label = f"{value_prefix}{true_prob:.3f}" if show_values else ""
    false_label = f"{value_prefix}{false_prob:.3f}" if show_values else ""

    # Create the HTML for a horizontal stacked bar
    html = f"""
```



```

<div style="width:100%; height:{height}; display:flex; border:1px solid #ccc; overflow:h
    <div style="flex-basis:{true_prob*100}%; background:linear-gradient(to bottom, rgba(
        <span style="font-size:10px; color:white; text-shadow:0px 0px 2px #000;">{true_l
    </div>
    <div style="flex-basis:{false_prob*100}%; background:linear-gradient(to bottom, rgba
        <span style="font-size:10px; color:white; text-shadow:0px 0px 2px #000;">{false_
    </div>
</div>
"""
return html

def create_tooltip(node_id, node_data):
    """
    Create rich HTML tooltip with probability information

    Args:
        node_id (str): Identifier of the node
        node_data (dict): Node attributes including probabilities

    Returns:
        str: HTML content for tooltip displayed on hover
    """
    # Extract node information
    description = node_data.get('description', '')
    priors = node_data.get('priors', {})
    instantiations = node_data.get('instantiations', ["TRUE", "FALSE"])

    # Start building the HTML tooltip
    html = f"""
    <div style="max-width:350px; padding:10px; background-color:#f8f9fa; border-radius:5px;
        <h3 style="margin-top:0; color:#202124;">{node_id}</h3>
        <p style="font-style:italic;">{description}</p>
    """

    # Add prior probabilities section
    if priors and 'true_prob' in priors:
        true_prob = priors['true_prob']
        false_prob = 1.0 - true_prob

        # Get proper state names
        true_state = instantiations[0] if len(instantiations) > 0 else "TRUE"
        false_state = instantiations[1] if len(instantiations) > 1 else "FALSE"

```

```

        html += f"""
        <div style="margin-top:10px; background-color:#fff; padding:8px; border-radius:4px; 1
            <h4 style="margin-top:0; font-size:14px;">Prior Probabilities:</h4>
            <div style="display:flex; justify-content:space-between; margin-bottom:4px;">
                <div style="font-size:12px;">{true_state}: {true_prob:.3f}</div>
                <div style="font-size:12px;">{false_state}: {false_prob:.3f}</div>
            </div>
            {create_probability_bar(true_prob, false_prob, "20px", True)}
        </div>
        """

    # Add click instruction
    html += """
    <div style="margin-top:8px; font-size:12px; color:#666; text-align:center;">
        Click node to see full probability details
    </div>
    </div>
    """

    return html

def create_expanded_content(node_id, node_data):
    """
    Create expanded content shown when a node is clicked

    Args:
        node_id (str): Identifier of the node
        node_data (dict): Node attributes including probabilities

    Returns:
        str: HTML content for detailed view displayed on click
    """
    # Extract node information
    description = node_data.get('description', '')
    priors = node_data.get('priors', {})
    posteriors = node_data.get('posteriors', {})
    instantiations = node_data.get('instantiations', ["TRUE", "FALSE"])

    # Get proper state names
    true_state = instantiations[0] if len(instantiations) > 0 else "TRUE"
    false_state = instantiations[1] if len(instantiations) > 1 else "FALSE"

```

```

# Extract probabilities
true_prob = priors.get('true_prob', 0.5)
false_prob = 1.0 - true_prob

# Start building the expanded content
html = f"""
<div style="max-width:500px; padding:15px; font-family:Arial, sans-serif;">
    <h2 style="margin-top:0; color:#333;">{node_id}</h2>
    <p style="font-style:italic; margin-bottom:15px;">{description}</p>

    <div style="margin-bottom:20px; padding:12px; border:1px solid #ddd; background-color:
        <h3 style="margin-top:0; color:#333;">Prior Probabilities</h3>
        <div style="display:flex; justify-content:space-between; margin-bottom:5px;">
            <div><strong>{true_state}</strong> {true_prob:.3f}</div>
            <div><strong>{false_state}</strong> {false_prob:.3f}</div>
        </div>
        {create_probability_bar(true_prob, false_prob, "25px", True)}
    </div>
    """"

# Add conditional probability table if available
if posteriors:
    html += """
    <div style="padding:12px; border:1px solid #ddd; background-color:#f9f9f9; border-ra
        <h3 style="margin-top:0; color:#333;">Conditional Probabilities</h3>
        <table style="width:100%; border-collapse:collapse; font-size:13px;">
            <tr style="background-color:#eee;">
                <th style="padding:8px; text-align:left; border:1px solid #ddd;">Conditio
                <th style="padding:8px; text-align:center; border:1px solid #ddd; width:
                <th style="padding:8px; text-align:center; border:1px solid #ddd;">Visual
            </tr>
        """"

    # Sort posteriors to group by similar conditions
    posterior_items = list(posteriors.items())
    posterior_items.sort(key=lambda x: x[0])

    # Add rows for conditional probabilities
    for key, value in posterior_items:
        try:
            # Try to parse probability value
            prob_value = float(value)

```

```

        inv_prob = 1.0 - prob_value

        # Add row with probability visualization
        html += f"""
        <tr>
            <td style="padding:8px; border:1px solid #ddd;">{key}</td>
            <td style="padding:8px; text-align:center; border:1px solid #ddd;">{prob}</td>
            <td style="padding:8px; border:1px solid #ddd;">
                {create_probability_bar(prob_value, inv_prob, "20px", False)}
            </td>
        </tr>
        """

    except:
        # Fallback for non-numeric values
        html += f"""
        <tr>
            <td style="padding:8px; border:1px solid #ddd;">{key}</td>
            <td style="padding:8px; text-align:center; border:1px solid #ddd;" colspan=
        </tr>
        """

    html += """
    </table>
</div>
"""

html += "</div>"

return html

```

Phase 4: Main Visualization Function

```

def create_bayesian_network_with_probabilities(df):
    """
    Create an interactive Bayesian network visualization with enhanced probability visualization
    and node classification based on network structure.
    """
    # Create a directed graph
    G = nx.DiGraph()

```

```

# Add nodes with proper attributes
for idx, row in df.iterrows():
    title = row['Title']
    description = row['Description']

    # Process probability information
    priors = get_priors(row)
    instantiations = get_instantiations(row)

    # Add node with base information
    G.add_node(
        title,
        description=description,
        priors=priors,
        instantiations=instantiations,
        posteriors=get_posteriors(row)
    )

# Add edges
for idx, row in df.iterrows():
    child = row['Title']
    parents = get_parents(row)

    # Add edges from each parent to this child
    for parent in parents:
        if parent in G.nodes():
            G.add_edge(parent, child)

# Classify nodes based on network structure
classify_nodes(G)

# Create network visualization
net = Network(notebook=True, directed=True, cdn_resources="in_line", height="600px", width="1000px")

# Configure physics for better layout
net.force_atlas_2based(gravity=-50, spring_length=100, spring_strength=0.02)
net.show_buttons(filter_=['physics'])

# Add the graph to the network
net.from_nx(G)

# Enhance node appearance with probability information and classification

```

```

for node in net.nodes:
    node_id = node['id']
    node_data = G.nodes[node_id]

    # Get node type and set border color
    node_type = node_data.get('node_type', 'unknown')
    border_color = get_border_color(node_type)

    # Get probability information
    priors = node_data.get('priors', {})
    true_prob = priors.get('true_prob', 0.5) if priors else 0.5

    # Get proper state names
    instantiations = node_data.get('instantiations', ["TRUE", "FALSE"])
    true_state = instantiations[0] if len(instantiations) > 0 else "TRUE"
    false_state = instantiations[1] if len(instantiations) > 1 else "FALSE"

    # Create background color based on probability
    background_color = get_probability_color(priors)

    # Create tooltip with probability information
    tooltip = create_tooltip(node_id, node_data)

    # Create a simpler node label with probability
    simple_label = f"{node_id}\np={true_prob:.2f}"

    # Store expanded content as a node attribute for use in click handler
    node_data['expanded_content'] = create_expanded_content(node_id, node_data)

    # Set node attributes
    node['title'] = tooltip # Tooltip HTML
    node['label'] = simple_label # Simple text label
    node['shape'] = 'box'
    node['color'] = {
        'background': background_color,
        'border': border_color,
        'highlight': {
            'background': background_color,
            'border': border_color
        }
    }
}

```

```

# Set up the click handler with proper data
setup_data = {
    'nodes_data': {node_id: {
        'expanded_content': json.dumps(G.nodes[node_id].get('expanded_content', '')),
        'description': G.nodes[node_id].get('description', ''),
        'priors': G.nodes[node_id].get('priors', {}),
        'posteriors': G.nodes[node_id].get('posteriors', {})
    } for node_id in G.nodes()}
}

# Add custom click handling JavaScript
click_js = """
// Store node data for click handling
var nodesData = %s;

// Add event listener for node clicks
network.on("click", function(params) {
    if (params.nodes.length > 0) {
        var nodeId = params.nodes[0];
        var nodeInfo = nodesData[nodeId];

        if (nodeInfo) {
            // Create a modal popup for expanded content
            var modal = document.createElement('div');
            modal.style.position = 'fixed';
            modal.style.left = '50%%';
            modal.style.top = '50%%';
            modal.style.transform = 'translate(-50%%, -50%%)';
            modal.style.backgroundColor = 'white';
            modal.style.padding = '20px';
            modal.style.borderRadius = '5px';
            modal.style.boxShadow = '0 0 10px rgba(0,0,0,0.5)';
            modal.style.zIndex = '1000';
            modal.style.maxWidth = '80%%';
            modal.style.maxHeight = '80%%';
            modal.style.overflow = 'auto';

            // Parse the JSON string back to HTML content
            try {
                var expandedContent = JSON.parse(nodeInfo.expanded_content);
                modal.innerHTML = expandedContent;
            } catch (e) {

```

```

        modal.innerHTML = 'Error displaying content: ' + e.message;
    }

    // Add close button
    var closeBtn = document.createElement('button');
    closeBtn.innerHTML = 'Close';
    closeBtn.style.marginTop = '10px';
    closeBtn.style.padding = '5px 10px';
    closeBtn.style.cursor = 'pointer';
    closeBtn.onclick = function() {
        document.body.removeChild(modal);
    };
    modal.appendChild(closeBtn);

    // Add modal to body
    document.body.appendChild(modal);
}
}
});
""" % json.dumps(setup_data['nodes_data'])

# Save the graph to HTML
html_file = "bayesian_network.html"
net.save_graph(html_file)

# Inject custom click handling into HTML
try:
    with open(html_file, "r") as f:
        html_content = f.read()

    # Insert click handling script before the closing body tag
    html_content = html_content.replace('</body>', f'<script>{click_js}</script></body>')

    # Write back the modified HTML
    with open(html_file, "w") as f:
        f.write(html_content)

    return HTML(html_content)
except Exception as e:
    return HTML(f"<p>Error rendering HTML: {str(e)}</p><p>The network visualization has l

```


Quickly check HTML Outputs

```
create_bayesian_network_with_probabilities(result_df)
```

```
# Use the function to create and display the visualization
```

```
print(result_df)
```

	Title \
0	Existential_Catastrophe
1	Human_Disempowerment
2	Scale_Of_Power_Seeking
3	Misaligned_Power_Seeking
4	APS_Systems
5	Advanced_AI_Capability
6	Agentic_Planning
7	Strategic_Awareness
8	Difficulty_Of_Alignment
9	Instrumental_Convergence
10	Problems_With_Proxies
11	Problems_With_Search
12	Deployment_Decisions
13	Incentives_To_Build_APS
14	Usefulness_Of_APS
15	Competitive_Dynamics
16	Deception_By_AI
17	Corrective_Feedback
18	Warning_Shots
19	Rapid_Capability_Escalation
20	Barriers_To_Understanding
21	Adversarial_Dynamics
22	Stakes_Of_Error

	Description	line	line_numbers \
0	The destruction of humanity's long-term potent...	0	[0]
1	Permanent and collective disempowerment of hum...	1	[1]
2	Power-seeking by AI systems scaling to the poi...	2	[2]
3	Deployed AI systems seeking power in unintende...	3	[3, 21, 23, 25]
4	AI systems with advanced capabilities, agentic...	4	[4]
5	AI systems that outperform humans on tasks tha...	5	[5]
6	AI systems making and executing plans based on...	6	[6]

7	AI systems with models accurately representing...	7	[7]
8	It is harder to build aligned systems than mis...	8	[8]
9	AI systems with misaligned objectives tend to ...	9	[9]
10	Optimizing for proxy objectives breaks correla...	10	[10]
11	Search processes can yield systems pursuing di...	11	[11]
12	Decisions to deploy potentially misaligned AI ...	12	[12]
13	Strong incentives to build and deploy APS syst...	13	[13]
14	APS systems are very useful for many valuable ...	14	[14]
15	Competitive pressures between AI developers.	15	[15]
16	AI systems deceiving humans about their true o...	16	[16]
17	Human society implementing corrections after o...	17	[17]
18	Observable failures in weaker systems before c...	18	[18]
19	AI capabilities escalating very rapidly, allow...	19	[19]
20	Difficulty in understanding the internal worki...	20	[20]
21	Potentially adversarial relationships between ...	22	[22]
22	The escalating impact of mistakes with power-s...	24	[24]

	indentation	indentation_levels	\
0	0		[0]
1	0		[0]
2	4		[4]
3	8	[8, 0, 0, 0]	
4	12		[12]
5	16		[16]
6	16		[16]
7	16		[16]
8	12		[12]
9	16		[16]
10	16		[16]
11	16		[16]
12	12		[12]
13	16		[16]
14	20		[20]
15	20		[20]
16	16		[16]
17	8		[8]
18	12		[12]
19	12		[12]
20	0		[0]
21	0		[0]
22	0		[0]

Parents \

```

0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22

```

[Scale_Of_Power_Seeking]

[Misaligned_Power_Seeking, Corrective_Feedback]

[APS_Systems, Difficulty_Of_Alignment, Deploym...]

[Advanced_AI_Capability, Agentic_Planning, Str...]

[Instrumental_Convergence, Problems_With_Proxi...]

[Incentives_To_Build_APS, Deception_By_AI]

[Usefulness_Of_APS, Competitive_Dynamics]

[Warning_Shots, Rapid_Capability_Escalation]

```

0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

```

Children \

[Human_Disempowerment]

[Scale_Of_Power_Seeking]

[Misaligned_Power_Seeking]

[APS_Systems]

[APS_Systems]

[APS_Systems]

[Misaligned_Power_Seeking]

[Difficulty_Of_Alignment]

[Difficulty_Of_Alignment]

[Difficulty_Of_Alignment]

[Misaligned_Power_Seeking]

[Deployment_Decisions]

[Incentives_To_Build_APS]

[Incentives_To_Build_APS]

[Deployment_Decisions]

[Scale_Of_Power_Seeking]

```

18 [Corrective_Feedback]
19 [Corrective_Feedback]
20 []
21 []
22 []

```

```

instantiations \
0 [existential_catastrophe_TRUE, existential_cat...
1 [human_disempowerment_TRUE, human_disempowerme...
2 [scale_of_power_seeking_TRUE, scale_of_power_s...
3 [misaligned_power_seeking_TRUE, misaligned_pow...
4 [aps_systems_TRUE, aps_systems_FALSE]
5 [advanced_ai_capability_TRUE, advanced_ai_capa...
6 [agentic_planning_TRUE, agentic_planning_FALSE]
7 [strategic_awareness_TRUE, strategic_awareness...
8 [difficulty_of_alignment_TRUE, difficulty_of_a...
9 [instrumental_convergence_TRUE, instrumental_c...
10 [problems_with_proxies_TRUE, problems_with_pro...
11 [problems_with_search_TRUE, problems_with_sear...
12 [deployment_decisions_DEPLOY, deployment_decis...
13 [incentives_to_build_aps_STRONG, incentives_to...
14 [usefulness_of_aps_HIGH, usefulness_of_aps_LOW]
15 [competitive_dynamics_STRONG, competitive_dyna...
16 [deception_by_ai_TRUE, deception_by_ai_FALSE]
17 [corrective_feedback_EFFECTIVE, corrective_fee...
18 [warning_shots_OBSERVED, warning_shots_UNOBSER...
19 [rapid_capability_escalation_TRUE, rapid_capab...
20 [barriers_to_understanding_HIGH, barriers_to_u...
21 [adversarial_dynamics_TRUE, adversarial_dynami...
22 [stakes_of_error_HIGH, stakes_of_error_LOW]

```

```

priors \
0 {'p(existential_catastrophe_TRUE)': '0.05', 'p...
1 {'p(human_disempowerment_TRUE)': '0.208', 'p(h...
2 {'p(scale_of_power_seeking_TRUE)': '0.208', 'p...
3 {'p(misaligned_power_seeking_TRUE)': '0.338', ...
4 {'p(aps_systems_TRUE)': '0.65', 'p(aps_systems...
5 {'p(advanced_ai_capability_TRUE)': '0.80', 'p(...
6 {'p(agentic_planning_TRUE)': '0.85', 'p(agenti...
7 {'p(strategic_awareness_TRUE)': '0.75', 'p(str...
8 {'p(difficulty_of_alignment_TRUE)': '0.40', 'p...
9 {'p(instrumental_convergence_TRUE)': '0.75', '...
10 {'p(problems_with_proxies_TRUE)': '0.80', 'p(p...

```

```

11 {'p(problems_with_search_TRUE)': '0.70', 'p(pr...
12 {'p(deployment_decisions_DEPLOY)': '0.70', 'p(...
13 {'p(incentives_to_build_aps_STRONG)': '0.80', ...
14 {'p(usefulness_of_aps_HIGH)': '0.85', 'p(usefu...
15 {'p(competitive_dynamics_STRONG)': '0.75', 'p(...
16 {'p(deception_by_ai_TRUE)': '0.50', 'p(decepti...
17 {'p(corrective_feedback_EFFECTIVE)': '0.60', '...
18 {'p(warning_shots_OBSERVED)': '0.70', 'p(warni...
19 {'p(rapid_capability_escalation_TRUE)': '0.45'...
20 {'p(barriers_to_understanding_HIGH)': '0.70', ...
21 {'p(adversarial_dynamics_TRUE)': '0.60', 'p(ad...
22 {'p(stakes_of_error_HIGH)': '0.85', 'p(stakes_...

```

	posteriors	No_Parent	No_Children	\
0	{'p(existential_catastrophe_TRUE human_disempo...	True	True	
1	{'p(human_disempowerment_TRUE scale_of_power_s...	False	True	
2	{'p(scale_of_power_seeking_TRUE misaligned_pow...	False	False	
3	{'p(misaligned_power_seeking_TRUE aps_systems_...	False	False	
4	{'p(aps_systems_TRUE advanced_ai_capability_TR...	False	False	
5	{}	True	False	
6	{}	True	False	
7	{}	True	False	
8	{'p(difficulty_of_alignment_TRUE instrumental_...	False	False	
9	{}	True	False	
10	{}	True	False	
11	{}	True	False	
12	{'p(deployment_decisions_DEPLOY incentives_to_...	False	False	
13	{'p(incentives_to_build_aps_STRONG usefulness_...	False	False	
14	{}	True	False	
15	{}	True	False	
16	{}	True	False	
17	{'p(corrective_feedback_EFFECTIVE warning_shot...	False	False	
18	{}	True	False	
19	{}	True	False	
20	{'p(barriers_to_understanding_HIGH misaligned_...	True	True	
21	{'p(adversarial_dynamics_TRUE misaligned_power...	True	True	
22	{'p(stakes_of_error_HIGH misaligned_power_seek...	True	True	

```

                                parent_instantiations
0                                []
1  [[scale_of_power_seeking_TRUE, scale_of_power_...
2  [[misaligned_power_seeking_TRUE, misaligned_po...
3  [[aps_systems_TRUE, aps_systems_FALSE], [diffi...

```

```

4  [[advanced_ai_capability_TRUE, advanced_ai_cap...
5                                     []
6                                     []
7                                     []
8  [[instrumental_convergence_TRUE, instrumental_...
9                                     []
10                                    []
11                                    []
12 [[incentives_to_build_aps_STRONG, incentives_t...
13 [[usefulness_of_aps_HIGH, usefulness_of_aps_LO...
14                                    []
15                                    []
16                                    []
17 [[warning_shots_OBSERVED, warning_shots_UNOBSE...
18                                    []
19                                    []
20                                    []
21                                    []
22                                    []

```

Conclusion: From Prototype to Production

Summary of Achievements

This notebook has successfully demonstrated the core AMTAIR extraction pipeline, transforming structured argument representations into interactive Bayesian network visualizations through the following steps:

1. **Environment Setup:** Established a reproducible environment with necessary libraries and data access
2. **Argument Extraction:** Processed structured ArgDown representations preserving the hierarchical relationships
3. **Probability Integration:** Enhanced arguments with probability information to create BayesDown
4. **Data Transformation:** Converted BayesDown into structured DataFrame representation
5. **Visualization & Analysis:** Created interactive Bayesian network visualizations with probability encoding

The rain-sprinkler-lawn example, though simple, demonstrates all the key components of the extraction pipeline that can be applied to more complex AI safety arguments.

Limitations and Future Work

While this prototype successfully demonstrates the core pipeline, several limitations and opportunities for future work remain:

1. **LLM Extraction:** The current implementation focuses on processing pre-formatted ArgDown rather than performing extraction directly from unstructured text. Future work will integrate LLM-powered extraction.
2. **Scalability:** The system has been tested on small examples; scaling to larger, more complex arguments will require additional optimization and handling of computational complexity.
3. **Policy Evaluation:** The current implementation focuses on representation and visualization; future work will add policy evaluation capabilities by implementing intervention modeling.
4. **Prediction Market Integration:** Future versions will integrate with forecasting platforms to incorporate live data into the models.

Connection to AMTAIR Project

This prototype represents just one component of the broader AMTAIR project described in the project documentation (see PY_AMTAIRDescription and PY_AMTAIR_SoftwareToolsNMilestones). The full project includes:

1. **AI Risk Pathway Analyzer (ARPA):** The core extraction and visualization system demonstrated in this notebook
2. **Worldview Comparator:** Tools for comparing different perspectives on AI risk
3. **Policy Impact Evaluator:** Systems for evaluating intervention effects across scenarios
4. **Strategic Intervention Generator:** Tools for identifying robust governance strategies

Together, these components aim to address the coordination crisis in AI governance by providing computational tools that make implicit models explicit, identify cruxes of disagreement, and evaluate policy impacts across diverse worldviews.

By transforming unstructured text into formal, analyzable representations, the AMTAIR project helps bridge the gaps between technical researchers, policy specialists, and other stakeholders, enabling more effective coordination in addressing existential risks from advanced AI.

6.0 Save Outputs

6. Saving and Exporting Results

This section provides tools for saving the notebook results and visualizations in various formats:

1. **HTML Export:** Creates a self-contained HTML version of the notebook with all visualizations
2. **Markdown Export:** Generates documentation-friendly Markdown version of the notebook
3. **PDF Export:** Creates a PDF document for formal sharing (requires LaTeX installation)

These exports are essential for: - Sharing analysis results with colleagues and stakeholders - Including visualizations in presentations and reports - Creating documentation for the AMTAIR project - Preserving results for future reference

The different formats serve different purposes, from interactive exploration (HTML) to documentation (Markdown) to formal presentation (PDF).

Instruction:

Download the ipynb, which you want to convert, on your local computer. Run the code below to upload the ipynb.

The html version will be downloaded automatically on your local machine. Enjoy it!

```
# @title 6.0 --- Save Visualization and Notebook Outputs as .HTML---

"""
BLOCK PURPOSE: Provides tools for saving the notebook results in various formats.

This block offers functions to:
1. Convert the notebook to HTML for easy sharing and viewing
2. Convert the notebook to Markdown for documentation purposes
3. Save the visualization outputs for external use

These tools are essential for preserving the analysis results and making them
accessible outside the notebook environment, supporting knowledge transfer
and integration with other AMTAIR project components.

DEPENDENCIES: nbformat, nbconvert modules
INPUTS: Current notebook state
OUTPUTS: HTML, Markdown, or other format versions of the notebook
```



```

"""

import nbformat
from nbconvert import HTMLExporter
import os

# Repository URL variable for file access
repo_url = "https://raw.githubusercontent.com/SingularitySmith/AMTAIR_Prototype/main/data/example_notebooks/rain-sprinkler-lawn.ipynb"
notebook_name = "AMTAIR_Prototype_example_carlsmith" # Change when working with different example notebooks

# Download the notebook file
!wget {repo_url}{notebook_name}.ipynb -O {notebook_name}.ipynb

# Load the notebook
try:
    with open(f"{notebook_name}.ipynb") as f:
        nb = nbformat.read(f, as_version=4)
        print(f" Successfully loaded notebook: {notebook_name}.ipynb")
except FileNotFoundError:
    print(f" Error: File '{notebook_name}.ipynb' not found. Please check if it was downloaded correctly.")

# Initialize the HTML exporter
exporter = HTMLExporter()

# Convert the notebook to HTML
try:
    (body, resources) = exporter.from_notebook_node(nb)

    # Save the HTML to a file
    with open(f"{notebook_name}IPYNB.html", "w") as f:
        f.write(body)
    print(f" Successfully saved HTML version to: {notebook_name}IPYNB.html")
except Exception as e:
    print(f" Error converting notebook to HTML: {str(e)}")

```

```

--2025-04-26 22:34:17-- https://raw.githubusercontent.com/SingularitySmith/AMTAIR_Prototype/main/data/example_notebooks/rain-sprinkler-lawn.ipynb
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.110.133, 185.199.108.133, 185.199.109.133, 185.199.111.133
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.110.133|:443... connected
HTTP request sent, awaiting response... 200 OK
Length: 1120047 (1.1M) [text/plain]
Saving to: 'AMTAIR_Prototype_example_rain-sprinkler-lawn.ipynb'

```

AMTAIR_Pr 0%[

]

0 --.-KB/s

AMTAIR_Prototy

2025-04-26 22:34:17 (16.4 MB/s) - 'AMTAIR_Prototype_example_rain-sprinkler-lawn.ipynb' saved

Successfully loaded notebook: AMTAIR_Prototype_example_rain-sprinkler-lawn.ipynb

Successfully saved HTML version to: AMTAIR_Prototype_example_rain-sprinkler-lawnIPYNB.html

Convert .ipynb Notebook to Markdown

```
# @title --- Convert .ipynb Notebook to Markdown ---

import nbformat
from nbconvert import MarkdownExporter
import os

# repo_url = "https://raw.githubusercontent.com/SingularitySmith/AMTAIR_Prototype/main/data/"
notebook_name = "AMTAIR_Prototype_example_carlsmith" #Change Notebook name and path when working

# Download the notebook file
!wget {repo_url}{notebook_name}.ipynb -O {notebook_name}.ipynb # Corrected line

# Load the notebook
# add error handling for file not found
try:
    with open(f"{notebook_name}.ipynb") as f:
        nb = nbformat.read(f, as_version=4)
except FileNotFoundError:
    print(f"Error: File '{notebook_name}.ipynb' not found. Please check if it was downloaded correctly")

# Initialize the Markdown exporter
exporter = MarkdownExporter(exclude_output=True) # Correct initialization

# Convert the notebook to Markdown
(body, resources) = exporter.from_notebook_node(nb)

# Save the Markdown to a file
with open(f"{notebook_name}IPYNB.md", "w") as f:
    f.write(body)
```

--2025-04-26 22:33:43-- https://raw.githubusercontent.com/SingularitySmith/AMTAIR_Prototype

Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.111.133, 185.199.111.133, 185.199.111.133, 185.199.111.133
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.111.133|:443...
HTTP request sent, awaiting response... 200 OK
Length: 1120047 (1.1M) [text/plain]
Saving to: 'AMTAIR_Prototype_example_rain-sprinkler-lawn.ipynb'

AMTAIR_Pr 0%[] 0 --.-KB/s AMTAIR_Protot

2025-04-26 22:33:43 (18.1 MB/s) - 'AMTAIR_Prototype_example_rain-sprinkler-lawn.ipynb' saved

```
# @title 6.1 --- Convert Notebook to Markdown Documentation ---

"""
BLOCK PURPOSE: Converts the notebook to Markdown format for documentation purposes.

Markdown is a lightweight markup language that is widely used for documentation
and is easily readable in both plain text and rendered formats. This conversion:

1. Preserves the structure and content of the notebook
2. Creates a format suitable for inclusion in documentation systems
3. Excludes code outputs to focus on the process and methodology
4. Supports version control and collaboration on GitHub

The resulting Markdown file can be used in project documentation, GitHub wikis,
or as a standalone reference guide to the AMTAIR extraction pipeline.

DEPENDENCIES: nbformat, nbconvert.MarkdownExporter modules
INPUTS: Current notebook state
OUTPUTS: Markdown version of the notebook
"""

import nbformat
from nbconvert import MarkdownExporter
import os

# Repository URL variable for file access
# repo_url = "https://raw.githubusercontent.com/SingularitySmith/AMTAIR_Prototype/main/data/"
notebook_name = "AMTAIR_Prototype_example_carlsmith" # Change when working with different ex

# Download the notebook file
!wget {repo_url}{notebook_name}.ipynb -O {notebook_name}.ipynb
```

```

# Load the notebook
try:
    with open(f"{notebook_name}.ipynb") as f:
        nb = nbformat.read(f, as_version=4)
        print(f" Successfully loaded notebook: {notebook_name}.ipynb")
except FileNotFoundError:
    print(f" Error: File '{notebook_name}.ipynb' not found. Please check if it was downloaded

# Initialize the Markdown exporter
exporter = MarkdownExporter(exclude_output=True) # Exclude outputs for cleaner documentation

# Convert the notebook to Markdown
try:
    (body, resources) = exporter.from_notebook_node(nb)

    # Save the Markdown to a file
    with open(f"{notebook_name}IPYNB.md", "w") as f:
        f.write(body)
    print(f" Successfully saved Markdown version to: {notebook_name}IPYNB.md")
except Exception as e:
    print(f" Error converting notebook to Markdown: {str(e)}")

```

```

--2025-04-26 22:31:45-- https://raw.githubusercontent.com/SingularitySmith/AMTAIR_Prototype
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.108.133, 185.199.
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.108.133|:443...
HTTP request sent, awaiting response... 200 OK
Length: 1120047 (1.1M) [text/plain]
Saving to: 'AMTAIR_Prototype_example_rain-sprinkler-lawn.ipynb'

```

```

AMTAIR_Pr  0%[                               ]      0  --.-KB/s      AMTAIR_Prototy

```

```

2025-04-26 22:31:45 (18.0 MB/s) - 'AMTAIR_Prototype_example_rain-sprinkler-lawn.ipynb' saved

```

```

Successfully loaded notebook: AMTAIR_Prototype_example_rain-sprinkler-lawn.ipynb
Successfully saved Markdown version to: AMTAIR_Prototype_example_rain-sprinkler-lawnIPYNB.m

```

```

import nbformat
from nbconvert import PDFExporter
import os
import subprocess

```

```

import re

def escape_latex_special_chars(text):
    """Escapes special LaTeX characters in a string."""
    latex_special_chars = ['&', '%', '#', '_', '{', '}', '~', '^', '\\']
    replacement_patterns = [
        (char, '\\' + char) for char in latex_special_chars
    ]

    # Escape reserved characters
    for original, replacement in replacement_patterns:
        text = text.replace(original, replacement) # This is the fix
    return text

# Function to check if a command is available
def is_command_available(command):
    try:
        subprocess.run([command], capture_output=True, check=True)
        return True
    except (subprocess.CalledProcessError, FileNotFoundError):
        return False

# Check if xelatex is installed, and install if necessary
if not is_command_available("xelatex"):
    print("Installing necessary TeX packages...")
    !apt-get install -y texlive-xetex texlive-fonts-recommended texlive-plain-generic
    print("TeX packages installed successfully.")
else:
    print("xelatex is already installed. Skipping installation.")

# repo_url = "https://raw.githubusercontent.com/SingularitySmith/AMTAIR_Prototype/main/data/"
notebook_name = "AMTAIR_Prototype_example_carlsmith" #Change Notebook name and path when working

# Download the notebook file
!wget {repo_url}{notebook_name}.ipynb -O {notebook_name}.ipynb # Corrected line

# Load the notebook
# add error handling for file not found
try:
    with open(f"{notebook_name}.ipynb") as f:
        nb = nbformat.read(f, as_version=4)
except FileNotFoundError:

```

```

print(f"Error: File '{notebook_name}.ipynb' not found. Please check if it was downloaded c

# Initialize the PDF exporter
exporter = PDFExporter(exclude_output=True) # Changed to PDFExporter

# Sanitize notebook cell titles to escape special LaTeX characters like '&'
for cell in nb.cells:
    if 'cell_type' in cell and cell['cell_type'] == 'markdown':
        if 'source' in cell and isinstance(cell['source'], str):
            # Replace '&' with '\protect&' in markdown cell titles AND CONTENT
            # Updated to use escape_latex_special_chars function
            cell['source'] = escape_latex_special_chars(cell['source'])
            # Additionally, escape special characters in headings
            cell['source'] = re.sub(r'(\#+)\$*(.*)', lambda m: m.group(1) + ' ' + escape_late

# Convert the notebook to PDF
(body, resources) = exporter.from_notebook_node(nb)

# Save the PDF to a file
with open(f"{notebook_name}IPYNB.pdf", "wb") as f: # Changed to 'wb' for binary writing
    f.write(body)

```

Installing necessary TeX packages...

Reading package lists... Done

Building dependency tree... Done

Reading state information... Done

The following additional packages will be installed:

```

dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono
fonts-texgyre fonts-urw-base35 libapache-pom-java libcommons-logging-java
libcommons-parent-java libfontbox-java libgs9 libgs9-common libidn12
libijs-0.35 libjbig2dec0 libkpathsea6 libpdfbox-java libptexenc1 libruby3.0
libsyntax2 libteckit0 libtexlua53 libtexluajit2 libwoff1 libzip-0-13
lmodern poppler-data preview-latex-style rake ruby ruby-net-telnet
ruby-rubygems ruby-webrick ruby-xmlrpc ruby3.0 rubygems-integration t1utils
teckit tex-common tex-gyre texlive-base texlive-binaries texlive-latex-base
texlive-latex-extra texlive-latex-recommended texlive-pictures tipa
xfonts-encodings xfonts-utils

```

Suggested packages:

```

fonts-noto fonts-freefont-otf | fonts-freefont-ttf libavalon-framework-java
libcommons-logging-java-doc libexcalibur-logkit-java liblog4j1.2-java
poppler-utils ghostscript fonts-japanese-mincho | fonts-ipafont-mincho
fonts-japanese-gothic | fonts-ipafont-gothic fonts-arphic-ukai
fonts-arphic-uming fonts-nanum ri ruby-dev bundler debhelper gv
| postscript-viewer perl-tk xpdf | pdf-viewer xzdec
texlive-fonts-recommended-doc texlive-latex-base-doc python3-pygments
icc-profiles libfile-which-perl libspreadsheet-parseexcel-perl
texlive-latex-extra-doc texlive-latex-recommended-doc texlive-luatex
texlive-pstricks dot2tex prerex texlive-pictures-doc vprerex
default-jre-headless tipa-doc

```

The following NEW packages will be installed:

```

dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono
fonts-texgyre fonts-urw-base35 libapache-pom-java libcommons-logging-java
libcommons-parent-java libfontbox-java libgs9 libgs9-common libidn12
libijs-0.35 libjbig2dec0 libkpathsea6 libpdfbox-java libptexenc1 libruby3.0
libsyntax2 libteckit0 libtexlua53 libtexlua53-luatex libwoff1 libzzip-0-13
lmodern poppler-data preview-latex-style rake ruby ruby-net-telnet
ruby-rubygems ruby-webrick ruby-xmlrpc ruby3.0 rubygems-integration t1utils
teckit tex-common tex-gyre texlive-base texlive-binaries
texlive-fonts-recommended texlive-latex-base texlive-latex-extra
texlive-latex-recommended texlive-pictures texlive-plain-generic
texlive-xetex tipa xfonts-encodings xfonts-utils

```

0 upgraded, 53 newly installed, 0 to remove and 34 not upgraded.

Need to get 182 MB of archives.

After this operation, 571 MB of additional disk space will be used.

```

Get:1 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-droid-fallback all 1:6.0.1r16-
Get:2 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-lato all 2.0-2.1 [2,696 kB]
Get:3 http://archive.ubuntu.com/ubuntu jammy/main amd64 poppler-data all 0.4.11-1 [2,171 kB]
Get:4 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-common all 6.17 [33.7 kB]
Get:5 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-urw-base35 all 20200910-1 [6,304 kB]
Get:6 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libgs9-common all 9.55.0~dfsg-0ubuntu1
Get:7 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libidn12 amd64 1.38-4ubuntu1
Get:8 http://archive.ubuntu.com/ubuntu jammy/main amd64 libijs-0.35 amd64 0.35-15build2 [16.7 kB]
Get:9 http://archive.ubuntu.com/ubuntu jammy/main amd64 libjbig2dec0 amd64 0.19-3build2 [64.7 kB]
Get:10 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libgs9 amd64 9.55.0~dfsg1-0ubuntu1
Get:11 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libkpathsea6 amd64 2021.2021.01-1
Get:12 http://archive.ubuntu.com/ubuntu jammy/main amd64 libwoff1 amd64 1.0.2-1build4 [45.2 kB]
Get:13 http://archive.ubuntu.com/ubuntu jammy/universe amd64 dvisvgm amd64 2.13.1-1 [1,221 kB]
Get:14 http://archive.ubuntu.com/ubuntu jammy/universe amd64 fonts-lmodern all 2.004.5-6.1 [4,700 kB]
Get:15 http://archive.ubuntu.com/ubuntu jammy/main amd64 fonts-noto-mono all 20210225-1build1
Get:16 http://archive.ubuntu.com/ubuntu jammy/universe amd64 fonts-texgyre all 20180621-3.1
Get:17 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libapache-pom-java all 18-1 [4,700 kB]

```

```

Get:18 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libcommons-parent-java all 43-1
Get:19 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libcommons-logging-java all 1.2
Get:20 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libptexenc1 amd64 2021.2021
Get:21 http://archive.ubuntu.com/ubuntu jammy/main amd64 rubygems-integration all 1.18 [5,33
Get:22 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 ruby3.0 amd64 3.0.2-7ubuntu
Get:23 http://archive.ubuntu.com/ubuntu jammy/main amd64 ruby-rubygems all 3.3.5-2 [228 kB]
Get:24 http://archive.ubuntu.com/ubuntu jammy/main amd64 ruby amd64 1:3.0~exp1 [5,100 B]
Get:25 http://archive.ubuntu.com/ubuntu jammy/main amd64 rake all 13.0.6-2 [61.7 kB]
Get:26 http://archive.ubuntu.com/ubuntu jammy/main amd64 ruby-net-telnet all 0.1.1-2 [12.6 k
Get:27 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 ruby-webrick all 1.7.0-3ubun
Get:28 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 ruby-xmlrpc all 0.3.2-1ubun
Get:29 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libruby3.0 amd64 3.0.2-7ubun
Get:30 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libsyntaxtex2 amd64 2021.2021
Get:31 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libteckit0 amd64 2.5.11+ds1-1 [
Get:32 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libtexlua53 amd64 2021.2021
Get:33 http://archive.ubuntu.com/ubuntu jammy-updates/main amd64 libtexluajit2 amd64 2021.20
Get:34 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libzzip-0-13 amd64 0.13.72+dfsg
Get:35 http://archive.ubuntu.com/ubuntu jammy/main amd64 xfonts-encodings all 1:1.0.5-0ubun
Get:36 http://archive.ubuntu.com/ubuntu jammy/main amd64 xfonts-utils amd64 1:7.7+6build2 [9
Get:37 http://archive.ubuntu.com/ubuntu jammy/universe amd64 lmodern all 2.004.5-6.1 [9,471
Get:38 http://archive.ubuntu.com/ubuntu jammy/universe amd64 preview-latex-style all 12.2-1ul
Get:39 http://archive.ubuntu.com/ubuntu jammy/main amd64 t1utils amd64 1.41-4build2 [61.3 kB]
Get:40 http://archive.ubuntu.com/ubuntu jammy/universe amd64 teckit amd64 2.5.11+ds1-1 [699
Get:41 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tex-gyre all 20180621-3.1 [6,20
Get:42 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 texlive-binaries amd64
Get:43 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-base all 2021.20220204-
Get:44 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-fonts-recommended all 2
Get:45 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-base all 2021.202
Get:46 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libfontbox-java all 1:1.8.16-2
Get:47 http://archive.ubuntu.com/ubuntu jammy/universe amd64 libpdfbox-java all 1:1.8.16-2 [
Get:48 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-recommended all 2
Get:49 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-pictures all 2021.20220
Get:50 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-latex-extra all 2021.20
Get:51 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-plain-generic all 2021.
Get:52 http://archive.ubuntu.com/ubuntu jammy/universe amd64 tipa all 2:1.3-21 [2,967 kB]
Get:53 http://archive.ubuntu.com/ubuntu jammy/universe amd64 texlive-xetex all 2021.20220204
Fetched 182 MB in 3s (69.8 MB/s)
Extracting templates from packages: 100%
Preconfiguring packages ...
Selecting previously unselected package fonts-droid-fallback.
(Reading database ... 126558 files and directories currently installed.)
Preparing to unpack .../00-fonts-droid-fallback_1%3a6.0.1r16-1.1build1_all.deb ...
Unpacking fonts-droid-fallback (1:6.0.1r16-1.1build1) ...

```


Selecting previously unselected package fonts-lato.
 Preparing to unpack .../01-fonts-lato_2.0-2.1_all.deb ...
 Unpacking fonts-lato (2.0-2.1) ...
 Selecting previously unselected package poppler-data.
 Preparing to unpack .../02-poppler-data_0.4.11-1_all.deb ...
 Unpacking poppler-data (0.4.11-1) ...
 Selecting previously unselected package tex-common.
 Preparing to unpack .../03-tex-common_6.17_all.deb ...
 Unpacking tex-common (6.17) ...
 Selecting previously unselected package fonts-urw-base35.
 Preparing to unpack .../04-fonts-urw-base35_20200910-1_all.deb ...
 Unpacking fonts-urw-base35 (20200910-1) ...
 Selecting previously unselected package libgs9-common.
 Preparing to unpack .../05-libgs9-common_9.55.0~dfsg1-0ubuntu5.11_all.deb ...
 Unpacking libgs9-common (9.55.0~dfsg1-0ubuntu5.11) ...
 Selecting previously unselected package libidn12:amd64.
 Preparing to unpack .../06-libidn12_1.38-4ubuntu1_amd64.deb ...
 Unpacking libidn12:amd64 (1.38-4ubuntu1) ...
 Selecting previously unselected package libijs-0.35:amd64.
 Preparing to unpack .../07-libijs-0.35_0.35-15build2_amd64.deb ...
 Unpacking libijs-0.35:amd64 (0.35-15build2) ...
 Selecting previously unselected package libjbig2dec0:amd64.
 Preparing to unpack .../08-libjbig2dec0_0.19-3build2_amd64.deb ...
 Unpacking libjbig2dec0:amd64 (0.19-3build2) ...
 Selecting previously unselected package libgs9:amd64.
 Preparing to unpack .../09-libgs9_9.55.0~dfsg1-0ubuntu5.11_amd64.deb ...
 Unpacking libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.11) ...
 Selecting previously unselected package libkpathsea6:amd64.
 Preparing to unpack .../10-libkpathsea6_2021.20210626.59705-1ubuntu0.2_amd64.deb ...
 Unpacking libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.2) ...
 Selecting previously unselected package libwoff1:amd64.
 Preparing to unpack .../11-libwoff1_1.0.2-1build4_amd64.deb ...
 Unpacking libwoff1:amd64 (1.0.2-1build4) ...
 Selecting previously unselected package dvisvgm.
 Preparing to unpack .../12-dvisvgm_2.13.1-1_amd64.deb ...
 Unpacking dvisvgm (2.13.1-1) ...
 Selecting previously unselected package fonts-lmodern.
 Preparing to unpack .../13-fonts-lmodern_2.004.5-6.1_all.deb ...
 Unpacking fonts-lmodern (2.004.5-6.1) ...
 Selecting previously unselected package fonts-noto-mono.
 Preparing to unpack .../14-fonts-noto-mono_20201225-1build1_all.deb ...
 Unpacking fonts-noto-mono (20201225-1build1) ...
 Selecting previously unselected package fonts-texgyre.

```

Preparing to unpack .../15-fonts-texgyre_20180621-3.1_all.deb ...
Unpacking fonts-texgyre (20180621-3.1) ...
Selecting previously unselected package libapache-pom-java.
Preparing to unpack .../16-libapache-pom-java_18-1_all.deb ...
Unpacking libapache-pom-java (18-1) ...
Selecting previously unselected package libcommons-parent-java.
Preparing to unpack .../17-libcommons-parent-java_43-1_all.deb ...
Unpacking libcommons-parent-java (43-1) ...
Selecting previously unselected package libcommons-logging-java.
Preparing to unpack .../18-libcommons-logging-java_1.2-2_all.deb ...
Unpacking libcommons-logging-java (1.2-2) ...
Selecting previously unselected package libptexenc1:amd64.
Preparing to unpack .../19-libptexenc1_2021.20210626.59705-1ubuntu0.2_amd64.deb ...
Unpacking libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package rubygems-integration.
Preparing to unpack .../20-rubygems-integration_1.18_all.deb ...
Unpacking rubygems-integration (1.18) ...
Selecting previously unselected package ruby3.0.
Preparing to unpack .../21-ruby3.0_3.0.2-7ubuntu2.10_amd64.deb ...
Unpacking ruby3.0 (3.0.2-7ubuntu2.10) ...
Selecting previously unselected package ruby-rubygems.
Preparing to unpack .../22-ruby-rubygems_3.3.5-2_all.deb ...
Unpacking ruby-rubygems (3.3.5-2) ...
Selecting previously unselected package ruby.
Preparing to unpack .../23-ruby_1%3a3.0~exp1_amd64.deb ...
Unpacking ruby (1:3.0~exp1) ...
Selecting previously unselected package rake.
Preparing to unpack .../24-rake_13.0.6-2_all.deb ...
Unpacking rake (13.0.6-2) ...
Selecting previously unselected package ruby-net-telnet.
Preparing to unpack .../25-ruby-net-telnet_0.1.1-2_all.deb ...
Unpacking ruby-net-telnet (0.1.1-2) ...
Selecting previously unselected package ruby-webrick.
Preparing to unpack .../26-ruby-webrick_1.7.0-3ubuntu0.1_all.deb ...
Unpacking ruby-webrick (1.7.0-3ubuntu0.1) ...
Selecting previously unselected package ruby-xmlrpc.
Preparing to unpack .../27-ruby-xmlrpc_0.3.2-1ubuntu0.1_all.deb ...
Unpacking ruby-xmlrpc (0.3.2-1ubuntu0.1) ...
Selecting previously unselected package libruby3.0:amd64.
Preparing to unpack .../28-libruby3.0_3.0.2-7ubuntu2.10_amd64.deb ...
Unpacking libruby3.0:amd64 (3.0.2-7ubuntu2.10) ...
Selecting previously unselected package libsyntax2:amd64.
Preparing to unpack .../29-libsyntax2_2021.20210626.59705-1ubuntu0.2_amd64.deb ...

```

```

Unpacking libsynctex2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libteckit0:amd64.
Preparing to unpack .../30-libteckit0_2.5.11+ds1-1_amd64.deb ...
Unpacking libteckit0:amd64 (2.5.11+ds1-1) ...
Selecting previously unselected package libtexlua53:amd64.
Preparing to unpack .../31-libtexlua53_2021.20210626.59705-1ubuntu0.2_amd64.deb ...
Unpacking libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libtexluajit2:amd64.
Preparing to unpack .../32-libtexluajit2_2021.20210626.59705-1ubuntu0.2_amd64.deb ...
Unpacking libtexluajit2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package libzip-0-13:amd64.
Preparing to unpack .../33-libzip-0-13_0.13.72+dfsg.1-1.1_amd64.deb ...
Unpacking libzip-0-13:amd64 (0.13.72+dfsg.1-1.1) ...
Selecting previously unselected package xfonts-encodings.
Preparing to unpack .../34-xfonts-encodings_1%3a1.0.5-0ubuntu2_all.deb ...
Unpacking xfonts-encodings (1:1.0.5-0ubuntu2) ...
Selecting previously unselected package xfonts-utils.
Preparing to unpack .../35-xfonts-utils_1%3a7.7+6build2_amd64.deb ...
Unpacking xfonts-utils (1:7.7+6build2) ...
Selecting previously unselected package lmodern.
Preparing to unpack .../36-lmodern_2.004.5-6.1_all.deb ...
Unpacking lmodern (2.004.5-6.1) ...
Selecting previously unselected package preview-latex-style.
Preparing to unpack .../37-preview-latex-style_12.2-1ubuntu1_all.deb ...
Unpacking preview-latex-style (12.2-1ubuntu1) ...
Selecting previously unselected package t1utils.
Preparing to unpack .../38-t1utils_1.41-4build2_amd64.deb ...
Unpacking t1utils (1.41-4build2) ...
Selecting previously unselected package teckit.
Preparing to unpack .../39-teckit_2.5.11+ds1-1_amd64.deb ...
Unpacking teckit (2.5.11+ds1-1) ...
Selecting previously unselected package tex-gyre.
Preparing to unpack .../40-tex-gyre_20180621-3.1_all.deb ...
Unpacking tex-gyre (20180621-3.1) ...
Selecting previously unselected package texlive-binaries.
Preparing to unpack .../41-texlive-binaries_2021.20210626.59705-1ubuntu0.2_amd64.deb ...
Unpacking texlive-binaries (2021.20210626.59705-1ubuntu0.2) ...
Selecting previously unselected package texlive-base.
Preparing to unpack .../42-texlive-base_2021.20220204-1_all.deb ...
Unpacking texlive-base (2021.20220204-1) ...
Selecting previously unselected package texlive-fonts-recommended.
Preparing to unpack .../43-texlive-fonts-recommended_2021.20220204-1_all.deb ...
Unpacking texlive-fonts-recommended (2021.20220204-1) ...

```

```

Selecting previously unselected package texlive-latex-base.
Preparing to unpack .../44-texlive-latex-base_2021.20220204-1_all.deb ...
Unpacking texlive-latex-base (2021.20220204-1) ...
Selecting previously unselected package libfontbox-java.
Preparing to unpack .../45-libfontbox-java_1%3a1.8.16-2_all.deb ...
Unpacking libfontbox-java (1:1.8.16-2) ...
Selecting previously unselected package libpdfbox-java.
Preparing to unpack .../46-libpdfbox-java_1%3a1.8.16-2_all.deb ...
Unpacking libpdfbox-java (1:1.8.16-2) ...
Selecting previously unselected package texlive-latex-recommended.
Preparing to unpack .../47-texlive-latex-recommended_2021.20220204-1_all.deb ...
Unpacking texlive-latex-recommended (2021.20220204-1) ...
Selecting previously unselected package texlive-pictures.
Preparing to unpack .../48-texlive-pictures_2021.20220204-1_all.deb ...
Unpacking texlive-pictures (2021.20220204-1) ...
Selecting previously unselected package texlive-latex-extra.
Preparing to unpack .../49-texlive-latex-extra_2021.20220204-1_all.deb ...
Unpacking texlive-latex-extra (2021.20220204-1) ...
Selecting previously unselected package texlive-plain-generic.
Preparing to unpack .../50-texlive-plain-generic_2021.20220204-1_all.deb ...
Unpacking texlive-plain-generic (2021.20220204-1) ...
Selecting previously unselected package tipa.
Preparing to unpack .../51-tipa_2%3a1.3-21_all.deb ...
Unpacking tipa (2:1.3-21) ...
Selecting previously unselected package texlive-xetex.
Preparing to unpack .../52-texlive-xetex_2021.20220204-1_all.deb ...
Unpacking texlive-xetex (2021.20220204-1) ...
Setting up fonts-lato (2.0-2.1) ...
Setting up fonts-noto-mono (20201225-1build1) ...
Setting up libwoff1:amd64 (1.0.2-1build4) ...
Setting up libtexlua53:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up libijs-0.35:amd64 (0.35-15build2) ...
Setting up libtexluaajit2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up libfontbox-java (1:1.8.16-2) ...
Setting up rubygems-integration (1.18) ...
Setting up libzip-0-13:amd64 (0.13.72+dfsg.1-1.1) ...
Setting up fonts-urw-base35 (20200910-1) ...
Setting up poppler-data (0.4.11-1) ...
Setting up tex-common (6.17) ...
update-language: texlive-base not installed and configured, doing nothing!
Setting up libjbig2dec0:amd64 (0.19-3build2) ...
Setting up libteckit0:amd64 (2.5.11+ds1-1) ...
Setting up libapache-pom-java (18-1) ...

```

```

Setting up ruby-net-telnet (0.1.1-2) ...
Setting up xfonts-encodings (1:1.0.5-0ubuntu2) ...
Setting up t1utils (1.41-4build2) ...
Setting up libidn12:amd64 (1.38-4ubuntu1) ...
Setting up fonts-texgyre (20180621-3.1) ...
Setting up libkpathsea6:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up ruby-webrick (1.7.0-3ubuntu0.1) ...
Setting up fonts-lmodern (2.004.5-6.1) ...
Setting up fonts-droid-fallback (1:6.0.1r16-1.1build1) ...
Setting up ruby-xmlrpc (0.3.2-1ubuntu0.1) ...
Setting up libsynchronet2:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up libgs9-common (9.55.0~dfsg1-0ubuntu5.11) ...
Setting up teckit (2.5.11+ds1-1) ...
Setting up libpdfbox-java (1:1.8.16-2) ...
Setting up libgs9:amd64 (9.55.0~dfsg1-0ubuntu5.11) ...
Setting up preview-latex-style (12.2-1ubuntu1) ...
Setting up libcommons-parent-java (43-1) ...
Setting up dvisvgm (2.13.1-1) ...
Setting up libcommons-logging-java (1.2-2) ...
Setting up xfonts-utils (1:7.7+6build2) ...
Setting up libptexenc1:amd64 (2021.20210626.59705-1ubuntu0.2) ...
Setting up texlive-binaries (2021.20210626.59705-1ubuntu0.2) ...
update-alternatives: using /usr/bin/xdvi-xaw to provide /usr/bin/xdvi.bin (xdvi.bin) in auto
update-alternatives: using /usr/bin/bibtex.original to provide /usr/bin/bibtex (bibtex) in auto
Setting up lmodern (2.004.5-6.1) ...
Setting up texlive-base (2021.20220204-1) ...
/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
/usr/bin/ucfr
mktexlsr: Updating /var/lib/texmf/ls-R-TEXLIVEDIST...
mktexlsr: Updating /var/lib/texmf/ls-R-TEXMFMAIN...
mktexlsr: Updating /var/lib/texmf/ls-R...
mktexlsr: Done.
tl-paper: setting paper size for dvips to a4: /var/lib/texmf/dvips/config/config-paper.ps
tl-paper: setting paper size for dvipdfmx to a4: /var/lib/texmf/dvipdfmx/dvipdfmx-paper.cfg
tl-paper: setting paper size for xdvi to a4: /var/lib/texmf/xdvi/XDvi-paper
tl-paper: setting paper size for pdftex to a4: /var/lib/texmf/tex/generic/tex-ini-files/pdftex
Setting up tex-gyre (20180621-3.1) ...
Setting up texlive-plain-generic (2021.20220204-1) ...
Setting up texlive-latex-base (2021.20220204-1) ...
Setting up texlive-latex-recommended (2021.20220204-1) ...
Setting up texlive-pictures (2021.20220204-1) ...

```

```

Setting up texlive-fonts-recommended (2021.20220204-1) ...
Setting up tipa (2:1.3-21) ...
Setting up texlive-latex-extra (2021.20220204-1) ...
Setting up texlive-xetex (2021.20220204-1) ...
Setting up rake (13.0.6-2) ...
Setting up libruby3.0:amd64 (3.0.2-7ubuntu2.10) ...
Setting up ruby3.0 (3.0.2-7ubuntu2.10) ...
Setting up ruby (1:3.0~exp1) ...
Setting up ruby-rubygems (3.3.5-2) ...
Processing triggers for man-db (2.10.2-1) ...
Processing triggers for mailcap (3.70+nmu1ubuntu1) ...
Processing triggers for fontconfig (2.13.1-4.2ubuntu5) ...
Processing triggers for libc-bin (2.35-0ubuntu3.8) ...
/sbin/ldconfig.real: /usr/local/lib/libhwloc.so.15 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_loader.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_5.so.3 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc.so.2 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbb.so.12 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbmalloc_proxy.so.2 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtcm.so.1 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_level_zero.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libur_adapter_openscl.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtcm_debug.so.1 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libumf.so.0 is not a symbolic link

/sbin/ldconfig.real: /usr/local/lib/libtbbbind_2_0.so.3 is not a symbolic link

Processing triggers for tex-common (6.17) ...
Running updmap-sys. This may take some time... done.
Running mktexlsr /var/lib/texmf ... done.
Building format(s) --all.

```

```
This may take some time... done.
TeX packages installed successfully.
--2025-04-26 22:32:56-- https://raw.githubusercontent.com/SingularitySmith/AMTAIR_Prototype
Resolving raw.githubusercontent.com (raw.githubusercontent.com)... 185.199.110.133, 185.199.
Connecting to raw.githubusercontent.com (raw.githubusercontent.com)|185.199.110.133|:443...
HTTP request sent, awaiting response... 200 OK
Length: 1120047 (1.1M) [text/plain]
Saving to: 'AMTAIR_Prototype_example_rain-sprinkler-lawn.ipynb'

AMTAIR_Prototype_ex 100%[=====>] 1.07M --.-KB/s in 0.06s

2025-04-26 22:32:56 (17.0 MB/s) - 'AMTAIR_Prototype_example_rain-sprinkler-lawn.ipynb' saved
```