Simulating the Impact of Targeted Health Information on Vaccine Acceptance: An LLM-Driven Agent-Based Modeling Plan

1. Introduction

Context

This report presents a comprehensive plan designed for a healthcare data analytics case competition. The central theme addresses the significant public health challenge of understanding and influencing vaccine decision-making. Vaccine acceptance is not static; it is shaped by a complex interplay of individual beliefs, social influences, and the information environment.[1] Particularly in the digital age, individuals navigate a flood of information, ranging from official public health guidance to pervasive misinformation, making it crucial to understand how different types of information impact choices. Traditional analytical methods, such as cross-sectional surveys, often capture only static snapshots of public opinion. Modeling the dynamic, iterative process through which individuals encounter information and update their beliefs necessitates more sophisticated methodologies capable of representing temporal dynamics and individual heterogeneity.[3]

Problem Statement

The core problem addressed by this plan is the need to effectively model and quantify the dynamic relationship between exposure to specific forms of public healthcare information—particularly news content with varying characteristics like factuality, tone, and framing—and an individual's evolving stance on vaccination. Existing research acknowledges the influence of information but often struggles to capture the granular, longitudinal nature of belief formation and change, especially as influenced by tailored or targeted information streams prevalent online.[5] There is a need for models that can simulate these micro-level processes to yield insights applicable to macro-level communication strategies.

Proposed Solution

To address this challenge, this plan outlines an innovative simulation framework grounded in Agent-Based Modeling (ABM) principles and powered by Large Language Models (LLMs). The simulation involves two interacting LLM agents operating within an iterative loop:

1. **Agent 1 (The 'Receiver'):** This agent simulates an individual user, embodying a specific profile encompassing demographic attributes, personality traits, and pre-existing beliefs

regarding vaccines.

2. **Agent 2 (The 'Influencer'):** This agent simulates an information source, dynamically generating news content tailored to the current state of the Receiver agent.

This approach capitalizes on the advanced capabilities of LLMs to simulate human-like reasoning, natural language interaction, and complex decision-making processes.[3] By leveraging LLMs, the simulation can achieve a higher degree of behavioral realism and representational richness compared to traditional ABM approaches that often rely on simplified rules and numerical state updates.[3] The iterative nature of the simulation allows for the observation and analysis of how the Receiver agent's vaccine acceptance evolves over time in response to the dynamically generated news from the Influencer agent. This methodology aligns with cutting-edge research in computational social science [15] and the burgeoning field of generative agent simulations.[7]

Objective

The objective of this report is to furnish a detailed, technically sound, and actionable plan for the design, implementation, and evaluation of this LLM-agent simulation. It provides comprehensive guidance on the conceptual underpinnings, agent design considerations (persona and news generation), data requirements and acquisition strategies, technical implementation using the specified toolkit (n8n on Azure, OpenAI, Google AI Studio), a rigorous evaluation framework, and critical ethical considerations pertinent to such a simulation.

Report Structure

The subsequent sections are organized as follows: Section 2 elaborates on the conceptual framework underpinning the simulation. Section 3 details the design specifications for the Receiver and Influencer LLM agents. Section 4 outlines the proposed implementation of the iterative simulation loop and the technical architecture using the designated tools. Section 5 identifies the necessary data sources and proposes acquisition strategies. Section 6 presents a multi-faceted strategy for evaluating the simulation's validity and outputs. Section 7 addresses potential challenges and delineates crucial ethical considerations. Finally, Section 8 concludes the report with a summary and actionable recommendations tailored for success in the healthcare data analytics competition.

2. Conceptual Framework: Simulating Information Influence on Vaccine Decisions

Theoretical Foundations

The proposed simulation integrates principles from several key domains:

- **Agent-Based Modeling (ABM):** ABM provides the foundational methodology. It is a computational approach that models systems by simulating the actions and interactions of autonomous entities, known as agents.[18] Instead of defining system-level equations, ABM focuses on defining agent attributes and behavioral rules at the micro-level. Complex macro-level patterns, such as shifts in collective opinion or behavior, emerge from these local interactions.[16] This "generative" approach, aiming to "grow" phenomena from the bottom up [16], is particularly well-suited for studying systems where individual heterogeneity, adaptation, and interaction are key drivers of system dynamics, contrasting with traditional top-down modeling techniques.[19] The focus is on local constructivity and agent autonomy driving system historicity.[18]

- **Computational Social Science (CSS):** This project falls within the domain of CSS, which employs computational methods to analyze and understand social phenomena.[17] The use of simulation, particularly ABM enhanced by AI, represents a significant trend in CSS, allowing researchers to explore complex social dynamics that are difficult to study using traditional empirical methods alone.[15] The integration of LLMs into ABM marks a notable advancement, creating "generative ABMs" capable of more nuanced simulations of human behavior.[3]
- **LLM-Driven Agents:** A pivotal element of this framework is the use of agents powered by LLMs. Traditional ABM often faces limitations in agent expressivity, relying on predefined rules or simplified state representations.[3] LLMs overcome many of these limitations by enabling agents to:
  - Process and generate information in natural language.[8]
  - Exhibit complex reasoning and decision-making based on internal states (beliefs, goals, persona) and external stimuli.[7]
  - Maintain memory of past interactions and information.[14]
  - Adapt behavior dynamically based on context.[7] This enhanced realism is crucial for modeling complex human decisions like vaccine acceptance, which are influenced by subtle linguistic cues, personal beliefs, and cognitive processes.[3] Research creating agents with simulated personalities [7], complex social behaviors [8], and specific applications like vaccine hesitancy modeling [3] underscore the potential of this approach.
- **Information Diffusion and Belief Dynamics:** The simulation directly models how information influences beliefs over time. It draws inspiration from research on influence diffusion in social networks [21] and the broader field of opinion dynamics.[13] These fields explore how individual opinions are formed and modified through communication and social interaction. The model incorporates factors identified as critical in shaping vaccine attitudes, including trust in information sources [1], perceived risk associated with disease and vaccines, the impact of misinformation [1], social influence [1], and the role of pre-existing beliefs and potential cognitive biases (like confirmation bias [13] or motivated reasoning [28]) in information processing.

Simulation Model Overview

The proposed model architecture centers around an iterative interaction loop between two LLM agents:

- **Agent 1 (Receiver):** This agent represents an individual from a target population. It is initialized with a specific persona, including demographic details, personality traits (e.g., based on the Big Five model), and initial beliefs and attitudes towards vaccination. Its primary function is to consume news content generated by Agent 2 and update its internal belief state accordingly.
- **Agent 2 (Influencer):** This agent acts as a dynamic information source, such as a personalized news feed or a simulated social media environment. Its role is to generate news snippets that are specifically targeted towards Agent 1, based on Agent 1's current state (e.g., its level of hesitancy, specific concerns, or personality traits).

The simulation proceeds through discrete time steps (iterations), following this core loop:

1. **Targeting:** Agent 2 receives information about Agent 1's current state.
2. **Generation:** Based on this input and a predefined communication strategy (e.g., type of framing, factuality level), Agent 2 generates a relevant news snippet using an LLM call.
3. **Consumption:** Agent 1 receives the generated news snippet.
4. **Update:** Agent 1 processes the news snippet through the lens of its persona and updates its internal state, particularly its vaccine acceptance level (represented numerically, e.g., on a Likert scale). This update involves an LLM call prompting the agent to reason and decide.
5. **Recording:** The state of Agent 1 (before and after the update), the generated news, and potentially the agent's reasoning are logged for analysis.
6. **Repetition:** The loop repeats for a predetermined number of iterations or until a specific condition is met (e.g., belief stabilization).

The primary goal is to observe and analyze the trajectory of Agent 1's vaccine acceptance score over these iterations. By systematically varying Agent 1's initial persona and Agent 2's news generation strategy, the simulation allows for testing hypotheses about the differential impact of various communication approaches on individuals with different characteristics.

Bridging Micro-Level Interaction with Macro-Level Communication Strategy

A key strength of this framework lies in its ability to connect the simulation of micro-level cognitive processes (how an individual agent updates its belief based on a single piece of information) with the macro-level goal of informing effective public health communication strategies.[31] ABM methodologies are inherently suited for exploring how complex system-level behaviors emerge from the aggregation of individual actions and interactions.[18]

The user's objective is not merely to model opinion change but to derive insights for *effective communication*. Effective communication often necessitates tailoring messages to resonate with specific audience segments, addressing their unique concerns, values, or psychological traits.[31] The proposed simulation directly facilitates this. By instantiating Agent 1 with diverse, data-grounded personas representing different population segments, and by programming Agent 2 to employ various communication strategies (e.g., different framings, tones, levels of factuality), the simulation becomes a testbed. It allows for exploring which types of messages (Agent 2's output) are most likely to positively influence (or negatively backfire with) specific

types of individuals (Agent 1's persona).
This aligns with research aiming to use simulations to estimate public opinion or discourse *before* deploying actual communication campaigns.[21] The simulation essentially functions as a dynamic, computational focus group, enabling the pre-evaluation and refinement of communication strategies targeted at specific population profiles. The outputs can offer concrete guidance on message design, moving beyond generic recommendations to suggest specific framings or arguments likely to be effective for individuals with particular demographic profiles, personality traits, or pre-existing beliefs.

3. Designing the LLM Agents

The success of the simulation hinges on the careful design and implementation of the two interacting LLM agents. Both agents require detailed specifications for their core components, generation methodologies, and mechanisms to ensure realism and consistency.

3.1 Agent 1: Simulating User Profiles (The 'Receiver')

This agent represents the individual whose vaccine decision-making process is being simulated. Its design must capture relevant individual differences that influence information processing and belief formation.

- **Persona Core Components:** The persona of Agent 1 needs to be multi-faceted to enable realistic simulation.[7] Key components include:
  - *Demographics:* Foundational attributes sampled from real-world population distributions, such as those provided by census data.[3] This includes age, gender, geographic region (can be simplified to urban/rural or state-level), and education level. Race/ethnicity can be incorporated if based on aggregated, anonymized data and handled with sensitivity to avoid stereotyping.
  - *Socioeconomic Status (SES):* Represented by income level or bracket and potentially broad occupational categories (e.g., healthcare worker, essential worker, office worker), as SES can correlate with health access, exposure risks, and potentially health beliefs.[33]
  - *Personality Traits:* Modeled using a standard framework like the Big Five Inventory (BFI) – Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism.[7] This allows for exploring personality-based differences in response to information and testing targeted messaging strategies.[32] Scores can be assigned based on sampling from known distributions or by defining specific personality archetypes relevant to health behaviors.
  - *Prior Beliefs & Attitudes:* This is a critical component defining the agent's starting point. It should include:
    - *Initial Vaccine Stance:* Quantified using a Likert-type scale (e.g., 1-4 or 1-5) representing the agent's initial willingness or hesitancy to accept the vaccine.[4]
    - *Trust Levels:* Degree of trust in key institutions like government health agencies (e.g., CDC), scientific bodies, pharmaceutical companies, and

media sources.[1]

        - *Specific Concerns/Narratives:* Susceptibility to or belief in common vaccine-related narratives, such as concerns about side effects, speed of development, efficacy, conspiracy theories, or specific misinformation themes.[32]
        - *Relevant Values:* Underlying values that might influence health decisions, such as emphasis on individual liberty, community well-being, natural immunity, or deference to authority.
    - *Cognitive Factors (Optional Extension):* For increased sophistication, one could attempt to explicitly model cognitive biases like confirmation bias (tendency to favor information confirming existing beliefs) [13] or motivated reasoning (processing information in a way that suits preferred conclusions).[28] However, this adds significant complexity. LLMs may already exhibit some implicit cognitive patterns based on their training data.[7]
- **Persona Generation Methodology:** Creating realistic and grounded personas requires a combination of data and synthesis:
    - *Data Grounding:* The foundation should be real-world data. Sample demographic profiles proportionally from census data.[3] Use public opinion surveys (e.g., GSS [7], Pew Research Center, Kaiser Family Foundation) to establish realistic distributions for initial vaccine stances, trust levels, and common concerns, linking these to demographics where data allows. Aggregate statistics from psychological research can inform personality trait distributions.
    - *LLM Synthesis:* Utilize the generative capabilities of OpenAI or Google AI Studio to weave the sampled parameters into a coherent and detailed natural language persona description or backstory for Agent 1.[14] This leverages the LLM's vast internal knowledge base, including embedded psychological and social science concepts [7], to create a richer, more plausible profile than just a list of parameters. The Stanford agent simulation study, which used extensive interviews to create rich backstories [7], provides a benchmark for the level of detail that enhances simulation accuracy, suggesting that even synthetically generated rich narratives are beneficial.
    - *Prompting Strategy:* Develop structured prompts for the LLM to generate these personas. The prompt should clearly define the required components (demographics, personality scores/descriptions, initial beliefs, values) and instruct the LLM to create a consistent profile. Strategies include handcrafting prompt templates filled with sampled data or providing the LLM with key parameters and asking it to generate the full profile narrative.[14] The prompt must explicitly state the agent's identity and its role within the simulation context.
- **Ensuring Realism and Consistency:** A major challenge with LLM agents is ensuring they consistently adhere to their assigned persona throughout the simulation.[9]
    - *Initial Validation:* Before running the main simulation loops, perform validation checks. Query the instantiated Agent 1 with questions related to its profile (e.g.,

hypothetical survey questions on attitudes, opinions, or even personality test items) and compare its responses to the expected answers based on its defined parameters and corresponding real-world data distributions.[7] The goal is to achieve a high degree of alignment, similar to the correlations observed in studies like the Stanford simulation (e.g., 85% GSS match, 80% personality correlation [7]).
  - *Consistency Mechanisms during Simulation:* Since LLMs can exhibit inconsistent values or drift from their roles over multiple interactions [3], implement mechanisms to reinforce persona fidelity within the iterative loop:
    - *Memory Integration:* Include the core persona description or key belief/personality statements within the prompt given to Agent 1 at each belief update step. This serves as a constant reminder of its identity and acts like a form of short-term memory.[14]
    - *Reflection Prompting:* Consider incorporating a reflection step, where the agent is prompted to briefly review its previous state, reasoning, or core persona elements before making its belief update decision. This can encourage more consistent decision-making.[9]
    - *Explicit Role Instruction:* Start the belief update prompt with clear instructions reinforcing the agent's role (e.g., "You are playing the role of. Based on this persona, evaluate the following news...").[14]
    - *Attitude Modulation/Warmup Adaptation:* If initial validation reveals significant discrepancies between agent responses and target distributions, consider adapting techniques like simulation warmup or attitude modulation mentioned in the VacSim framework.[3] This might involve adjusting prompt parameters or running preliminary iterations to better align the agent's baseline responses before the main experimental simulation.
- **The Persona Fidelity Paradox:** A critical consideration emerges from conflicting research findings. On one hand, studies demonstrate that LLMs can simulate individual personalities derived from rich data (like interviews) with impressive accuracy.[7] On the other hand, research also highlights significant limitations: LLMs may struggle to maintain distinct personalities consistently, sometimes reverting to an average persona or exhibiting contradictory values across different contexts.[9] Some studies found that persona variables accounted for less than 10% of the variance in simulated responses [9], and inconsistencies between agent responses and their demographic profiles have been noted in simulations like VacSim.[3]
  This apparent paradox suggests that simply providing an LLM with a persona description might not be sufficient for robust simulation. High-fidelity simulation likely requires a two-pronged approach:
  1. **Rich Input:** Grounding the persona in detailed information, whether real (like interviews [7]) or synthetically generated rich backstories based on sampled data, seems crucial for achieving initial accuracy.[7] Simple demographic prompts alone may be insufficient.[9]

2. **Active Enforcement:** The simulation design must actively work to *maintain* the persona's integrity throughout the iterative process. This necessitates incorporating mechanisms like persistent memory cues in prompts, reflection steps, or explicit instructions to constantly reinforce the agent's assigned role and characteristics.[9] Persona generation is therefore not a one-time setup but requires ongoing management within the simulation loop. Continuous validation or consistency checks during the simulation are advisable.[3]

3.2 Agent 2: Generating Targeted News (The 'Influencer')

This agent simulates the information environment by dynamically generating news content tailored to influence or interact with Agent 1.

- **News Content Core Components:** The generated news snippets must have controllable attributes to allow for systematic experimentation:
  - *Topic Focus:* Primarily centered on the relevant vaccine (e.g., COVID-19 vaccines, potentially others like HPV or influenza for different scenarios [36]).
  - *Factuality/Stance:* Classified according to established categories used in misinformation research and fact-checking:
    - Factual/True: Accurate information, often pro-vaccine or neutral reporting.[37]
    - Misinformation/Fake/False: Demonstrably false claims, potentially anti-vaccine.[37]
    - Misleading: Contains elements of truth but presented out of context, altered, or with critical omissions.[40]
    - Opinion/Commentary: Expresses a viewpoint rather than stating facts.[42]
    - Satire: Intended as humor, but potentially misconstrued.[40]
    - (Consider also Disinformation: Deliberately false with intent to deceive [43]; Mal-information: True information used to harm [43]).
  - *Tone/Framing:* Control over the rhetorical and emotional style of the news snippet:
    - Neutral/Objective: Reportorial style.
    - Emotional: Fear-based (e.g., emphasizing side effects, risks [2]), Hope-based (e.g., emphasizing efficacy, return to normalcy), Anger-inducing, Empathetic.
    - Authoritative: Citing experts or official bodies.
    - Anecdotal: Using personal stories or testimonials.
    - Scientific/Technical: Using complex language or data.
    - Conspiracy-focused: Hinting at hidden agendas or plots.[36]
  - *Source Attribution (Simulated):* Assign a simulated source to the news snippet to potentially influence Agent 1's perception of credibility.[27] Levels could include: High Credibility (e.g., CDC, WHO, Major Medical Journal), Medium Credibility (e.g., Mainstream News Outlet), Low Credibility (e.g., Known Misinformation Website, Anonymous Blog), Unverifiable (e.g., Social Media Post).
- **News Generation Methodology:**

    - *LLM-Powered Synthetic Data Generation:* Utilize OpenAI/Google AI Studio as the primary engine for generating news snippets. Employ sophisticated prompt engineering to control the output based on the desired parameters (topic, factuality, tone, simulated source).[45] Crucially, the prompt must incorporate information about Agent 1's current state (e.g., belief score, key personality traits, expressed concerns) to ensure the generated news is *targeted*. Techniques like model distillation (using a larger model to generate data) could be explored if optimizing for cost/speed becomes necessary, but likely sticking to primary models is sufficient.[46] Frameworks like DataDreamer [50] or concepts from SynthLLM [54] offer structured approaches to synthetic data generation, providing conceptual guidance.
    - *Grounding with Real Examples:* To ensure the generated news feels authentic, ground the generation process in real-world data. Use examples from curated datasets of actual and fake news related to vaccines [37] as few-shot examples within the generation prompt. Alternatively, analyze these datasets to extract characteristic linguistic features, topics, or rhetorical patterns [42] that can be specified in the prompt to guide the LLM's style and content.
    - *Targeting Mechanism:* This is fundamental to the simulation's premise. The generation prompt for Agent 2 must be dynamically constructed in each iteration to include relevant aspects of Agent 1's current state. For example, the prompt might specify: "Generate a short, misleading news article (missing context) about vaccine efficacy, using an authoritative tone, targeted at an individual who is highly conscientious and currently expresses 'probably yes' (score 3/4) vaccine intention."
    - *Iterative Refinement (Recommended):* To improve the quality and consistency of generated news, consider adding a self-refinement step.[35] After Agent 2 generates an initial news snippet, use a subsequent LLM prompt to ask the model (or another LLM) to critique the snippet based on criteria like clarity, persuasiveness, adherence to the specified factuality/tone, and relevance to the target profile. Then, use another prompt to instruct the LLM to revise the snippet based on its own critique. This feedback loop can help produce more polished and precisely controlled stimuli for Agent 1.
- **Ensuring Content Quality and Relevance:** Validating the output of Agent 2 is essential.
    - *Automated Checks:* Implement basic checks using simpler models or rules (e.g., sentiment analysis for tone [42], keyword checks for topic relevance). Consider using an LLM-as-a-judge approach [9], where a separate LLM call evaluates the generated snippet against the intended parameters (e.g., "Rate the factuality of the following text on a scale of 1-5"). Automated classifiers trained on real/fake news data could also be used.[42]
    - *Manual Review:* No automated check is perfect. Periodically sample the news snippets generated by Agent 2 and manually review them for coherence,

    plausibility, relevance to the target state of Agent 1, and accurate reflection of the intended factuality, tone, and source characteristics.
  - *Diversity Control:* Monitor the output to prevent the generator from producing overly repetitive or simplistic content. Techniques include varying the generation prompts, adjusting LLM temperature settings, using diverse seed examples from real-world data [41], and potentially employing techniques specifically designed for diverse synthetic data generation.[45]
- **Dynamic vs. Static News Generation:** The requirement for *targeted* news within an *iterative* simulation necessitates dynamic generation. Static news datasets [41], while useful for grounding, cannot adapt to the evolving state of the Receiver agent (Agent 1). The simulation aims to model how beliefs change in response to an information stream that is potentially reactive or tailored. For the news generated by Agent 2 to be truly "targeted," its content must be contingent on Agent 1's current beliefs, concerns, or other relevant persona attributes. A static pool of articles would lack this crucial responsiveness. Therefore, the news generation process cannot be pre-computed before the simulation starts. It must be an integral part of each simulation iteration, with Agent 2's LLM call dynamically generating content based on the state of Agent 1 from the previous step. This creates a closed loop where Agent 1's beliefs influence the information it receives, which in turn influences its subsequent beliefs, mirroring real-world feedback dynamics in personalized information environments. The technical implementation must explicitly link the output of Agent 1's belief update to the input for Agent 2's news generation in the next cycle.

Table 1: LLM Agent Design Parameters

Parameter Category	**Agent 1 (Receiver) Parameters**	**Agent 2 (Influencer) Parameters**	**Data Sources / Generation Method**
Core Identity	Demographics (Age, Gender, Education, Region, SES)	Simulated Source Type (e.g., Health Authority, News Outlet, Blog)	Census Data, LLM Synthesis
	Personality (Big Five Traits: O, C, E, A, N)	Source Credibility Level (High, Medium, Low, Unverifiable)	Literature Norms, LLM Synthesis / Assigned
Beliefs / Content	Initial Vaccine Stance (Likert Scale 1-4)	News Factuality/Stance (Fact, Misinfo, Misleading, Opinion)	Survey Data (Pew, KFF, GSS), LLM Synthesis / Controlled Parameter
	Trust Levels (Science, Govt, Media)	News Tone/Framing (Neutral, Emotional, Authoritative, Anecdotal)	Survey Data / Controlled Parameter
	Specific Concerns/Narratives (Side effects,	News Topic Focus (Vaccine efficacy, safety, mandates, etc.)	Survey Data, Literature / Controlled Parameter

	Conspiracy belief)		
Behavior / Function	Belief Update Mechanism (LLM reasoning based on persona+news)	News Generation Mechanism (LLM generates targeted snippet)	LLM (OpenAI/Google AI Studio) Prompting
	Output: Updated Belief Score (1-4) + Justification Text	Input: Agent 1 State (Score, Persona Snippets)	LLM Prompting / Workflow Logic
Consistency	Memory Integration (Persona in prompt), Reflection (Optional)	Grounding (Seed examples from real news), Refinement (Optional)	Prompt Engineering, Workflow Design / Real News Datasets [41], Self-Refine [35]
Validation	Initial checks vs. Surveys, Local Consistency Checks	Automated Checks (LLM Judge, Classifier), Manual Review	GSS [7], Survey Data / LLM-as-Judge [9], Classifiers [42], Manual Sampling

4. Implementing the Iterative Simulation Loop

This section details the step-by-step logic of the simulation, the method for quantifying belief change, the proposed technical architecture using n8n, OpenAI, and Google AI Studio, and considerations regarding specialized agent frameworks.

4.1 Core Simulation Logic

The simulation operates in discrete time steps, often referred to as 'ticks' in ABM literature.[23] The process involves an initialization phase followed by an iterative loop.

- **Initialization Phase (Tick 0):**
    1. *Define Agent 1 Persona:* Sample or specify the parameters for Agent 1's persona (demographics, personality traits, initial beliefs, trust levels) based on the distributions derived from grounding data (Section 5).
    2. *Generate Persona Description:* Use an LLM call (OpenAI/Google AI Studio) with a structured prompt containing the sampled parameters to generate a coherent natural language description and backstory for Agent 1. Store this description as part of the agent's persistent state.
    3. *Set Initial Belief State:* Assign Agent 1's starting vaccine acceptance score on the chosen Likert scale (e.g., 1-4).
    4. *Define Agent 2 Strategy:* Specify the communication strategy Agent 2 will employ for this particular simulation run (e.g., "Consistently generate factual news addressing common side effect concerns," or "Generate fear-mongering misinformation targeting individuals with high neuroticism scores"). Store these strategy parameters.
- **Iterative Phase (Ticks 1 to N):** This loop represents the core dynamic of the simulation.

For each tick:

1. *Prepare Input for Agent 2:* Extract relevant information from Agent 1's current state. This must include the current belief score and potentially key persona elements relevant to Agent 2's strategy (e.g., specific fears identified in the persona, trust level, relevant personality traits). Format this information according to Agent 2's prompt template.
2. *Generate News (Agent 2):* Execute an LLM API call (OpenAI/Google AI Studio) using the prepared prompt. The LLM generates a news snippet consistent with the specified strategy and targeted based on Agent 1's state.
3. *Log Generated News:* Store the generated news snippet along with the current tick number.
4. *Prepare Input for Agent 1:* Construct the prompt for Agent 1's belief update. This prompt must include:
    - The full persona description (or relevant summary) to maintain context and consistency.
    - Agent 1's belief score from the *previous* tick.
    - The news snippet generated in step 2.
    - Clear instructions for the task (reassess belief, provide score, provide justification).
5. *Update Belief (Agent 1):* Execute an LLM API call (OpenAI/Google AI Studio) using Agent 1's prompt. The LLM simulates the agent's reasoning process based on its persona and the news, outputting the new belief score and a textual justification.
6. *Update State & Log:* Parse the LLM response to extract the new numerical belief score and the textual justification. Update Agent 1's state variables. Log the following information for this tick: tick number, Agent 1's belief score *before* the update, the received news snippet, the generated justification, and Agent 1's belief score *after* the update.
7. *Check Termination Condition:* Determine if the simulation should continue. This could be based on reaching a predefined maximum number of iterations (N) or meeting a convergence criterion (e.g., Agent 1's belief score remaining stable for a certain number of consecutive ticks). If the condition is not met, increment the tick counter and return to step 1 of the iterative phase.

4.2 Quantifying Belief Change

A quantitative measure of Agent 1's belief state is essential for tracking influence over time.

- **Measurement Scale:** A Likert-type scale is a practical and commonly used method for representing attitudes or intentions in surveys and simulations.[26] A 4-point scale, as used in VacSim [4], offers a clear distinction without a neutral midpoint:
    - 1 = Definitely No (Strong Hesitancy/Refusal)
    - 2 = Probably No (Moderate Hesitancy)
    - 3 = Probably Yes (Moderate Acceptance)
    - 4 = Definitely Yes (Strong Acceptance) This scale provides an ordinal measure of

vaccine acceptance/intention.

- **Belief Update Mechanism:** The core of the belief change process lies in the LLM call for Agent 1 during the update step (Step 5 in the iterative phase). The prompt design is critical. It must explicitly instruct the LLM to:
  1. Adopt the specified persona of Agent 1.
  2. Consider its current belief state (score from the previous tick).
  3. Process the specific news snippet it just received.
  4. Output a revised belief score on the defined 1-4 scale based on its simulated reasoning. The prompt should constrain the output format to ensure the numerical score can be easily extracted.
- **Tracking and Analysis:** The primary quantitative data generated by the simulation is the time series of Agent 1's belief scores across iterations. Analyzing this series allows for quantifying the impact of different news strategies (Agent 2) on different personas (Agent 1). Key quantitative metrics include:
  - *Overall Change:* Difference between the final and initial belief score.
  - *Direction:* Whether the belief score trended towards increased acceptance (positive change) or increased hesitancy (negative change).
  - *Rate of Change:* The average change in belief score per iteration, indicating the speed of influence.
  - *Volatility:* Measures like the standard deviation of belief scores over time, indicating stability or fluctuation in beliefs. These metrics can be compared statistically across different experimental conditions (variations in Agent 1 persona and Agent 2 strategy).
- **Beyond Scalar Beliefs - Capturing Reasoning:** Relying solely on a numerical belief score, while necessary for quantitative tracking, misses crucial information about *why* an agent's belief changed. LLMs possess the unique ability to articulate their reasoning process in natural language [3], unlike traditional ABM agents that typically follow predefined mathematical update rules.[13] Capturing this reasoning provides invaluable qualitative insights for understanding the mechanisms of influence and designing effective communication [User Query].

  To achieve this, the belief update prompt for Agent 1 should be modified to request not only the new numerical score but also a brief textual justification for that score. For example: "...Respond with your new score (1-4) followed by a concise (1-2 sentence) explanation of why you arrived at this score, considering the news you read and your personal profile." This generated text should be logged alongside the numerical score at each iteration. Analyzing this qualitative data (e.g., through thematic analysis or content coding) can reveal:
  - Which specific arguments or pieces of information in the news were most salient to the agent.
  - How the agent's persona (e.g., trust level, personality, prior concerns) mediated its interpretation of the news.
  - Why certain messages were persuasive and others were not, or even counterproductive. This qualitative dimension significantly enhances the

explanatory power of the simulation, providing actionable insights for crafting communication strategies that resonate with specific audience concerns and reasoning patterns. The evaluation plan (Section 6) must incorporate methods for analyzing this rich textual data.

4.3 Technical Architecture (n8n, OpenAI, Google AI Studio)

The simulation loop can be implemented using the specified tools, with n8n serving as the central orchestrator.

- **Orchestration (n8n on Azure):** The n8n workflow will automate the sequence described in Section 4.1. Key n8n nodes and their functions include:
  - *Manual Trigger / Cron Node:* To start a simulation run.
  - *Set Node / Function Node:* To initialize simulation parameters (persona details, Agent 2 strategy, max iterations) and agent state variables at Tick 0.
  - *HTTP Request Node / Dedicated AI Nodes (e.g., OpenAI Node, Google Vertex AI Node):* To execute the necessary API calls to OpenAI and/or Google AI Studio for:
    - Generating the initial persona description (Tick 0).
    - Generating the targeted news snippet by Agent 2 in each iteration.
    - Performing the belief update (score + justification) for Agent 1 in each iteration.
  - *Function Node / Code Node (JavaScript or Python):* To handle data manipulation, such as:
    - Parsing LLM responses to extract the numerical score and textual justification.
    - Updating Agent 1's state variables based on the LLM output.
    - Dynamically constructing the prompts for the next LLM calls based on the current state and parameters.
    - Formatting data for logging.
  - *Looping Implementation:* Use n8n's control flow nodes. A common pattern involves:
    - Using a SplitInBatches Node (set to batch size 1) combined with a loop back to itself, controlled by a counter or an IF node checking the termination condition.
    - Alternatively, structure the workflow such that the output of the logging step feeds back into the input of the next iteration's news generation step, using IF nodes to manage the exit condition.
  - *Data Logging:* Use nodes to write the logged data from each iteration to a persistent storage solution on Azure (e.g., Azure Blob Storage for CSV/JSON files, Azure SQL Database, or Azure Table Storage) or potentially Google Sheets for simpler cases.
- **LLM Integration:** Securely store API keys for OpenAI and Google AI Studio using n8n's credential management. Select appropriate LLM models (e.g., GPT-4/GPT-4o, Gemini models) based on the specific requirements of each task (persona generation, news generation, belief update/reasoning) considering factors like capability, cost, and rate

limits. Implement error handling for API calls.

- **State Management in n8n:** This is a critical implementation detail. Since n8n workflows don't inherently maintain state across separate executions or easily across complex loops involving multiple API calls, Agent 1's evolving state (belief score, potentially a summary of recent interactions or justifications) needs careful management. Two primary approaches:
    1. *Pass State via Data Flow:* Structure the workflow so that the output data item from one iteration contains all necessary state information required for the next iteration. This data item is then passed as input to the subsequent iteration's nodes. This is feasible for a two-agent model but can become complex if the state grows large.
    2. *External State Storage:* Use an external database (e.g., Azure SQL, Azure Table Storage) or even a simple file storage (Azure Blob) to store Agent 1's state. At the beginning of each iteration, an n8n node reads the current state; at the end, another node writes the updated state back. This is generally more robust and scalable, especially for longer simulations or more complex states. **This external storage approach is recommended for better reliability.**
- **Tool Synergy and Limitations:** The combination of n8n for orchestration and OpenAI/Google AI Studio for LLM capabilities provides a direct way to implement the simulation using the specified tools. n8n excels at connecting APIs and automating sequential workflows [66], making it suitable for executing the defined iterative steps. However, it's important to recognize that n8n is a general-purpose workflow automation tool, not a specialized framework for agent-based modeling or multi-agent systems like Mesa [20], AutoGen [14], or CrewAI.[69] These dedicated frameworks often provide built-in abstractions for managing agent state, memory, scheduling, and complex interaction patterns, which might need to be implemented manually within the n8n workflow logic. While the proposed two-agent loop is manageable in n8n, particularly with external state storage, scaling the simulation to include many interacting agents or more complex social network dynamics would likely be significantly more challenging and less efficient within n8n compared to using a dedicated ABM/MAS framework. The choice of n8n prioritizes adherence to the user's specified toolkit but entails careful design for state management and acknowledges potential scalability limits.

Table 2: Iterative Simulation Workflow in n8n (Conceptual)

Step	n8n Node Type(s)	Function	Input Data	Output Data	State Management
0. Init	Set / Function / HTTP Request	Define parameters, Generate Agent 1 persona description, Set initial belief	Simulation config (persona params, strategy)	Initial Agent 1 state (persona desc, score), Strategy params	Write initial state to external storage (e.g., Azure DB)

		score			
Loop Start (Tick t)	Read Storage Node / Function	Retrieve Agent 1's state from previous tick (t-1)	Tick number (t)	Agent 1 state (score, relevant persona snippets)	Read state from external storage
1. Prep News Gen Input	Function / Set	Format Agent 1 state & strategy params for Agent 2 prompt	Agent 1 state, Strategy params	Prompt input for Agent 2	-
2. Generate News (Agent 2)	HTTP Request / AI Node	Call LLM API to generate targeted news snippet	Prompt input for Agent 2	Generated news snippet	-
3. Prep Belief Update Input	Function / Set	Format Agent 1 persona, previous score, & generated news for Agent 1 prompt	Agent 1 persona desc, Agent 1 score (t-1), Generated news	Prompt input for Agent 1	-
4. Update Belief (Agent 1)	HTTP Request / AI Node	Call LLM API to get new belief score & justification	Prompt input for Agent 1	New belief score (t), Justification text	-
5. Update & Log State	Function / Write Storage Node	Parse LLM output, Update Agent 1 score, Log tick data (t, score(t-1), news, justification, score(t))	New score (t), Justification, Data from previous steps (news, score(t-1))	Updated Agent 1 state (score(t)), Log entry	Write updated state to external storage
6. Check Termination	IF / Function	Check if max iterations reached or convergence met	Tick number (t), Agent 1 score history (optional)	Boolean (continue/stop)	-
Loop End	Merge / No Operation / End Workflow	If continue, loop back to "Loop Start"; If stop, end workflow	Output from Step 6	Final logs / simulation results	-

4.4 Agent Framework Considerations (AutoGen, CrewAI, etc.)

While the primary implementation relies on n8n, awareness of specialized LLM agent frameworks can inform the design and offer alternatives if complexity increases.

- **Context:** Frameworks like LangChain [24], LlamaIndex [24], AutoGen [14], CrewAI [69], and others provide structured environments for building applications with LLM agents. They often include reusable components for common agent functions such as planning (breaking down tasks), memory (short-term context, long-term knowledge retrieval), and tool use (interacting with external APIs or data sources).[14]
- **Potential Utility & Comparison:**
    - *Component Libraries (LangChain, LlamaIndex):* These offer modules that could potentially simplify parts of the agent logic if implemented in Python scripts callable by n8n. For instance, LangChain's memory modules could offer more sophisticated ways to manage Agent 1's history than simple prompt injection.
    - *Multi-Agent Frameworks (AutoGen, CrewAI):* These are designed specifically for orchestrating interactions between multiple agents.
        - **CrewAI:** Emphasizes role-based collaboration within structured workflows.[69] Its paradigm maps naturally to the 'Receiver' and 'Influencer' roles defined in this project. It is often considered to have a lower barrier to entry for automating known processes.[73]
        - **AutoGen:** Offers greater flexibility in defining conversational patterns and agent interactions, making it suitable for more dynamic or complex problem-solving where the interaction flow might not be strictly predefined.[14] It generally requires more coding proficiency.[73]
- **Recommendation for this Project:** Given the user's specified toolkit, the primary implementation should remain within n8n, leveraging OpenAI and Google AI Studio APIs directly. The concepts and structures from agent frameworks should serve as *conceptual guides* for designing the prompts and workflow logic within n8n. For instance, explicitly managing short-term context (recent news) and long-term context (persona) in Agent 1's prompts mimics memory management.[14] The logic determining Agent 2's output based on Agent 1's state acts as a simple planning mechanism.[14] Direct integration of these frameworks should only be considered a fallback option if insurmountable challenges arise in managing state or interaction complexity purely within the n8n environment.

5. Data Requirements and Acquisition

Grounding the simulation in realistic data is crucial for its validity and relevance. This requires acquiring data for both initializing the Receiver agent's persona and guiding the Influencer agent's news generation.

Data for Agent 1 (Receiver) Persona Grounding

Creating diverse and representative personas for Agent 1 requires integrating data from multiple sources:

- **Demographics:**
    - *Source:* National statistical agencies (e.g., US Census Bureau for US-based simulations).
    - *Data:* Publicly available aggregated data tables providing distributions for age groups, gender, educational attainment, income brackets, and potentially geographic indicators (e.g., urban/rural, state/region). Cross-tabulations are valuable if available (e.g., education by age group).
    - *Use:* To sample realistic demographic profiles for Agent 1 instances, ensuring the simulated population reflects real-world diversity.[3]
    - *Access:* Typically downloadable directly from government websites.
- **Personality (Big Five Traits):**
    - *Source:* Academic psychological research papers, meta-analyses, or large-scale survey reports that publish population norms or distributions for Big Five scores (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism).
    - *Data:* Mean scores, standard deviations, and potential correlations with demographic variables for the target population. Direct datasets with individual scores are often not public due to privacy.
    - *Use:* To assign plausible personality scores or profiles to Agent 1, enabling the exploration of personality-based responses to information.[32] Can involve sampling from estimated distributions or defining archetypes (e.g., "high neuroticism, low agreeableness").
    - *Access:* Requires literature search and potentially synthesizing information from multiple published sources. Alternatively, rely on the LLM's implicit knowledge by prompting for specific personality types.
- **Beliefs/Attitudes:**
    - *Source:* Reputable public opinion survey organizations (e.g., Pew Research Center, Kaiser Family Foundation (KFF), Gallup) and academic surveys (e.g., General Social Survey (GSS) [7], European Social Survey (ESS) [28]).
    - *Data:* Survey questions and responses regarding:
        - Vaccine confidence/hesitancy/intention (ideally using scales similar to the one planned for the simulation).
        - Trust in institutions (government, health authorities, scientists, media).
        - Specific reasons for hesitancy (e.g., concerns about side effects, efficacy,

speed of development, ingredients).
    - Belief in specific misinformation narratives or conspiracy theories.
    - Political ideology or party affiliation (often correlates with health beliefs). Look for surveys that provide breakdowns by demographic groups.
  - *Use:* To set realistic initial belief states (vaccine acceptance score) and associated attitudes (trust levels, specific concerns) for Agent 1, ensuring the starting points reflect real-world opinion distributions.
  - *Access:* Many survey results are publicly available on organizational websites. Accessing microdata (individual-level responses) might require registration or specific data use agreements.

Data for Agent 2 (Influencer) News Generation

Guiding Agent 2 to produce relevant and realistic news content requires examples and potentially structured information:

- **News Content Examples (Seed Data):**
  - *Source:* Curated datasets specifically focused on health (mis)information, particularly related to vaccines/COVID-19. Examples include:
    - CoAID [41]: Contains fake news, user engagement data (check specific contents and availability).
    - CoVID19-FNIR [37]: Includes fact-checked fake news (from Poynter) and true news (from verified Twitter handles) related to COVID-19 from US, India, Europe (Feb-June 2020).
    - Li et al. (2021) dataset [38]: COVID-19 rumors from news/tweets with sentiment and stance labels.
    - Patwa et al. (2020) dataset [39]: Manually annotated dataset of real and fake news on COVID-19 from social media and articles. Also consider scraping examples from major news outlets (e.g., Reuters, BBC, NYT [66]) and known misinformation sources identified by fact-checkers.[80]
  - *Data:* Text content of news articles, social media posts, or headlines, ideally labeled with factuality (real/fake/misleading), stance, or sentiment.
  - *Use:* As seed examples (few-shot learning) in the prompts for Agent 2's LLM to guide its generation style, tone, and content themes.[82] Can also be used to extract linguistic features [42] or topics associated with different types of news to inform prompt design. Fine-tuning a dedicated generator model is a more advanced option.
  - *Access:* Public repositories like Kaggle [41] or IEEE DataPort [37], potentially direct downloads from research paper repositories, or web scraping (subject to ethical considerations).
- **Fact-Checking Data:**
  - *Source:* Established fact-checking organizations (e.g., Snopes [80], PolitiFact [80], FactCheck.org [80], Reuters Fact Check [87]) and potentially aggregated sources via

APIs like the Google Fact Check Tools API.[90]
    - *Data:* Collections of claims paired with veracity ratings (e.g., True, False, Misleading, Pants-on-Fire [86]), along with explanations or debunks.
    - *Use:* To provide ground truth examples for different factuality levels required by Agent 2. Can be used as examples in prompts, to potentially train an auxiliary classifier to evaluate Agent 2's output, or queried via API to source claims or check generated content.
    - *Access:* Primarily through websites (scraping may be necessary, check terms) or the Google Fact Check Tools API [90], which appears to be the most viable option for programmatic access to a broad range of fact-checks. Direct public APIs for individual organizations like Snopes or PolitiFact seem unavailable for general use.[40] Meta's API relates to their platform content [40], and Reuters offers commercial API access.[87]

Data Acquisition Strategies

Multiple methods will likely be needed:

- **Direct Download/Access:** Prioritize obtaining data directly from official sources (Census, survey archives, established dataset repositories like Kaggle, IEEE DataPort, university archives).
- **API Integration:** Leverage APIs where available, primarily the Google Fact Check Tools API [90] for accessing fact-check data programmatically within the n8n workflow. This requires obtaining and managing an API key.[90]
- **Web Scraping (Use Ethically and Judiciously):** For collecting news examples or fact-checks from websites that do not offer public APIs.
    - *Tools:* Utilize standard Python libraries like requests, BeautifulSoup [78], Scrapy [78], or Selenium for websites heavily reliant on JavaScript.[78]
    - *Ethical Guidelines:* **Strict adherence to ethical scraping practices is mandatory.** Always check and respect the website's robots.txt file. Implement appropriate delays between requests (rate limiting) to avoid overloading servers. Use a descriptive User-Agent string to identify the scraper's purpose. Handle errors gracefully. Abide by the website's terms of service.[78] Failure to do so can lead to IP blocks or legal issues.
    - *Managed Services:* Consider using commercial scraping services like Firecrawl [78] or Scrapfly [66] (which offers n8n integration). These services often handle complexities like proxy rotation, CAPTCHA solving, and browser fingerprinting, but typically involve costs.

The Data Synthesis Imperative

A crucial realization is that a single, perfect dataset containing linked information on individual demographics, personality, specific vaccine beliefs, and longitudinal news exposure is highly unlikely to exist publicly. Available data sources are fragmented – demographic data is

separate from opinion surveys, which are separate from news archives and fact-check databases.[7]
Therefore, this project necessitates not just data *acquisition* but significant data *synthesis* and *integration*. This involves:

1. Combining demographic data [10] with survey data on beliefs and attitudes [7] to create grounded initial profiles for Agent 1.
2. Potentially using LLMs themselves to fill gaps in these profiles or generate coherent persona narratives based on the disparate data points [14], leveraging their ability to synthesize information.
3. Using examples of real-world news and misinformation [41] not just as static inputs, but as guiding examples or stylistic targets for the *synthetic generation* of dynamic, targeted news content by Agent 2.[45]

This reliance on synthesis means that the LLMs specified in the toolkit (OpenAI, Google AI Studio) will likely play a role in the data preparation pipeline itself, in addition to their core function within the simulation loop. This adds a layer of methodological complexity but is essential for achieving the depth and dynamism required by the simulation's goals. Careful documentation of the synthesis process is vital for transparency and reproducibility.

Table 3: Potential Datasets and Data Sources

Data Type	**Potential Source(s)**	**Description / Relevance**	**Access Method**
Demographics	US Census Bureau (or relevant national agency)	Aggregated distributions (age, gender, education, income, region)	Public Download
Personality (Big Five)	Academic Literature / Survey Reports	Population norms, distributions, correlations with demographics	Literature Search
Vaccine Beliefs/Attitudes	Pew Research, KFF, Gallup, GSS, ESS	Surveys on vaccine confidence, trust, hesitancy reasons, political lean	Public Reports/Datasets
Real/Fake News Examples	CoAID [41], CoVID19-FNIR [37], Li et al. [38], Patwa et al. [39] Datasets	Labeled examples of vaccine/COVID-19 news/misinformation	Download (check sources)
	News Outlets (Reuters, BBC, NYT), Misinformation Sites	Unlabeled examples for style/topic grounding	Web Scraping (Ethical)
Fact-Checking Data	Google Fact Check	Searchable database of fact-checked claims	API (Requires Key)

	Tools API [90]	from multiple sources	
	Snopes, PolitiFact, FactCheck.org Websites	Specific examples of claims, ratings, and debunks	Web Scraping (Ethical)

6. Evaluation and Analysis Strategy

A rigorous evaluation plan is essential to assess the simulation's credibility, understand its behavior, and derive meaningful insights for the competition. The evaluation should address the simulation's validity, the realism of the agents' behavior, and the effectiveness of different communication strategies, aligning with goals established in similar simulation studies like VacSim.[3]

Evaluation Goals

1. **Assess Simulation Validity:** Determine if the simulation framework behaves plausibly and consistently.
2. **Evaluate Agent Realism:** Check if the LLM agents act in ways consistent with their defined personas and human-like behavior.
3. **Extract Communication Insights:** Quantify and explain how different news generation strategies impact belief change for various user profiles.

Metrics

Metrics should cover different levels of analysis:

- **Simulation Validity Metrics:**
    - *Real-World Alignment (Macro-Level):* Compare aggregate trends observed in the simulation (e.g., average belief shift under specific information conditions) with real-world data, if available and comparable (e.g., shifts in public opinion polls on vaccine confidence during periods with known information campaigns or events).[3] This is challenging but provides crucial external validation.
    - *Global Consistency (Internal):* Verify that the simulation responds predictably to changes in key parameters.[3] For example:
        - Does increasing the proportion of negative/misleading news generated by Agent 2 lead to a greater decrease in vaccine acceptance for susceptible Agent 1 personas?
        - Does increasing the simulated credibility of factual news sources lead to greater positive belief shifts? Run sensitivity analyses by systematically varying parameters and observing outcome trends.
- **Agent Realism Metrics:**
    - *Local Consistency (Micro-Level):* Evaluate whether Agent 1's individual belief updates and justifications are logically consistent with its defined persona (beliefs, personality, trust levels) and the specific news item it received in that iteration.[3] This often requires qualitative review of logged justifications, looking for

contradictions or out-of-character responses. Be mindful of known LLM inconsistencies.[3]
    - *Persona Accuracy (Micro-Level):* As part of initial validation (Section 3.1), and potentially periodically during the simulation, assess how well Agent 1's responses to probe questions (e.g., simulated survey items from GSS, BFI, or attitude scales) match the expected responses based on its profile.[7] Quantify using correlation or accuracy measures if ground truth is available.
- **Communication Effectiveness Metrics:**
    - *Differential Belief Change:* Measure the difference in the final belief score (or the total change over N iterations) for identical Agent 1 personas exposed to different Agent 2 news generation strategies.
    - *Influence Magnitude:* Quantify the average belief shift per news exposure under different conditions.
    - *Identify Influential Factors:* Determine which news characteristics (factuality, tone, source) and which persona characteristics (initial belief, trust, personality) are most strongly associated with significant belief changes (positive or negative).

Analysis Methods

A mixed-methods approach combining quantitative and qualitative analysis is recommended:

- **Quantitative Analysis:**
    - *Descriptive Statistics:* Calculate means, medians, standard deviations [22], and ranges for belief scores at different time points and across different conditions.
    - *Inferential Statistics:* Use statistical tests (e.g., t-tests, ANOVA, regression analysis) to compare belief trajectories and final scores between different experimental groups (e.g., persona type A vs. persona type B under misinformation strategy X).
    - *Time Series Analysis:* Analyze the patterns of belief change over iterations (e.g., identifying trends, convergence points, oscillations).
    - *Performance Metrics:* If evaluating internal classifications (e.g., Agent 2's adherence to factuality parameters), use standard metrics like Accuracy, Precision, Recall, and F1-Score.[42] If comparing simulation output to a target trajectory or real data, metrics like Mean Squared Error (MSE) or R^2 can be useful.[22]
- **Qualitative Analysis:**
    - *Thematic Analysis/Content Coding:* Analyze the textual justifications logged during Agent 1's belief update step [Insight 4.1]. Identify recurring themes, reasoning patterns, arguments cited, and emotional responses expressed by the agent in relation to different news types and persona characteristics. This provides crucial context for the quantitative changes observed. Software for qualitative data analysis (QDAS) can assist if the volume of text is large.
    - *Case Studies:* Select representative simulation runs (e.g., a highly hesitant persona exposed to factual news, a trusting persona exposed to misinformation)

and perform in-depth qualitative analysis of the interaction log to illustrate key dynamics.

- **Visualization:** Effective visualization is key for understanding and communicating results:
  - *Line Plots:* Show the trajectory of Agent 1's belief score over iterations for individual runs or averaged across conditions. Compare trajectories for different personas/strategies on the same plot.
  - *Bar Charts / Box Plots:* Compare final belief scores or total belief change across different experimental conditions.
  - *Histograms / Density Plots:* Show the distribution of final belief scores for different groups.
  - *Heatmaps:* Visualize correlations between persona traits, news characteristics, and belief change magnitude.
  - *Tools:* Standard Python libraries like Matplotlib [23], Seaborn, and potentially Plotly for interactive plots. R packages like ggplot2 are also powerful options.[22] While direct use of ABM visualization tools like Mesa's SolaraViz [20] isn't planned, their principles (interactive exploration, agent state display) can inspire custom visualizations if needed. Network visualization tools (e.g., NetworkX [23]) would be relevant if scaling to multiple interacting agents.

The Need for Multi-Method Evaluation

Evaluating a complex simulation involving nuanced LLM behavior requires a comprehensive strategy that goes beyond simple outcome metrics or internal consistency checks. Research involving LLM agents and social simulations emphasizes the importance of validating against real-world data or human behavior where possible [3], checking for consistency at both the individual agent (local) and system (global) levels [10], and employing both quantitative [22] and qualitative methods.[33]

Given the simulation's goal of understanding the *process* of information influence and the known complexities and potential inconsistencies of LLMs [9], a single metric like the final belief score is insufficient for rigorous validation. A robust evaluation must therefore integrate multiple perspectives:

1. **Quantitative Checks:** Assessing internal consistency (does the simulation behave predictably when parameters change?) [10] and, where feasible, external validity (do aggregate patterns align with real-world trends?).[3]
2. **Agent-Level Assessment:** Evaluating local consistency (does the agent's reasoning align with its persona and the immediate context?) [10] and persona fidelity (does the agent accurately represent its profile?).[7]
3. **Process-Level Understanding:** Utilizing qualitative analysis of the agents' generated reasoning [Insight 4.1] to understand *why* changes occur, adding depth and explanatory power to the quantitative findings. This multi-faceted approach ensures a more thorough and credible assessment of the simulation's behavior and the reliability of its

insights.

Table 4: Evaluation Metrics and Methods

Evaluation Goal	Metric Category	Specific Metric(s)	Analysis Method(s)	Data Source(s)
Simulation Validity	Real-World Alignment	Correlation/Comparison of simulated aggregate trends vs. real-world survey data	Statistical Comparison (if data permits), Qualitative Comparison	Simulation Logs, Public Opinion Poll Data
	Global Consistency	Predictable changes in outcomes (e.g., avg. belief score) when parameters vary	Sensitivity Analysis, Regression, ANOVA	Simulation Logs (multiple runs w/ varied params)
Agent Realism	Local Consistency	Qualitative assessment of reasoning alignment with persona & context	Manual Review of Justifications, Thematic Analysis	Simulation Logs (Justification Text)
	Persona Accuracy	Agent responses to probe questions vs. expected/ground truth	Correlation, Accuracy Scores, Comparison to survey distributions	Agent Responses to Probes, Survey Data [7]
Communication Effectiveness	Differential Impact	Difference in final belief score / total change between strategies for same persona	T-tests, ANOVA, Effect Size Calculation	Simulation Logs (Belief Scores)
	Influence Magnitude	Average belief shift per iteration/exposure under different conditions	Descriptive Statistics, Regression	Simulation Logs (Belief Scores)
	Key Factor Identification	Association between news/persona features and belief change	Correlation Analysis, Regression Modeling, Qualitative	Simulation Logs (All logged variables)

		magnitude/direction	Analysis of Justifications	

7. Addressing Challenges and Ethical Considerations

Implementing and interpreting this LLM-agent simulation involves navigating both technical hurdles and significant ethical responsibilities.

Technical Challenges & Mitigation Strategies

- **LLM Consistency and Reliability:** LLMs are known to sometimes produce inconsistent outputs, deviate from instructions, or "hallucinate" information.[3] This can affect both the persona adherence of Agent 1 and the controlled generation of news by Agent 2.
  - *Mitigation:* Employ careful and explicit prompt engineering with clear instructions and constraints.[97] Use techniques like few-shot prompting with good examples. Consider implementing self-correction or refinement loops (e.g., Self-Refine [35]) where the LLM critiques and improves its own output before it's used in the simulation. Monitor agent behavior for inconsistencies during runs (local consistency checks). Select LLMs known for better instruction following and stability; some studies suggest variations in performance across models for simulation tasks.[11]
- **Simulation Fidelity:** Ensuring the simplified two-agent model captures relevant aspects of the complex real-world process of information influence and vaccine decision-making is challenging.
  - *Mitigation:* Ground agent personas and news generation parameters in empirical data as much as possible.[7] Validate simulation outputs against real-world trends or patterns where feasible.[3] Be transparent about the model's assumptions (e.g., isolated two-agent interaction vs. complex social network [8]), simplifications, and limitations in the final report.[98] Acknowledge that this model explores a specific mechanism in isolation.
- **Scalability:** The current plan focuses on a two-agent interaction. While insightful, it doesn't capture broader social dynamics. Scaling to a larger population of interacting agents would significantly increase computational costs (due to numerous LLM API calls per tick) and workflow management complexity within n8n.
  - *Mitigation:* For the competition, maintain the focus on the two-agent model, which is computationally feasible and allows for deep analysis of the targeted influence mechanism. Clearly state scalability to larger populations as an area for future work, potentially referencing more scalable ABM frameworks [8] or LLM optimization techniques (e.g., using smaller/cheaper models for certain tasks, batching API calls if possible).
- **Quantifying Belief:** Representing a complex psychological construct like vaccine acceptance or hesitancy with a single Likert scale score is inherently a simplification.
  - *Mitigation:* Acknowledge this limitation explicitly. Crucially, supplement the

quantitative score by capturing and analyzing the agent's qualitative reasoning [Insight 4.1]. This provides a richer understanding that mitigates the reductionism of the numerical scale alone.

Ethical Considerations

Simulating sensitive topics like vaccine hesitancy and potentially generating misinformation requires careful ethical deliberation and adherence to established guidelines.

- **Misinformation Generation and Amplification:** Agent 2 is designed to potentially generate news content classified as misinformation or misleading for experimental purposes. There is a risk, however small, that the simulation could inadvertently generate novel harmful narratives or that the process itself could be seen as contributing to the problem.[36]
    - *Mitigation:* Exercise strict control over Agent 2's generation process. Use prompts that rely heavily on grounded examples from vetted real-world misinformation datasets [41] rather than allowing completely free-form generation of false content. Implement content filters or manual review steps for generated news, especially if exploring highly sensitive or potentially harmful themes. Frame the research ethically, emphasizing the goal of *understanding* influence to *combat* misinformation. Adhere strictly to the principle of non-maleficence (do no harm).[98]
- **Bias Representation and Stereotyping:** Personas based on demographic data carry a risk of reflecting or even amplifying societal biases and stereotypes if not constructed and used carefully. LLMs themselves can inherit biases from their training data.[14]
    - *Mitigation:* Use aggregated, anonymized demographic data for grounding. Ensure representative sampling across demographic groups relevant to the research question. Critically evaluate all generated persona descriptions for potentially biased or stereotypical portrayals before use. Test the simulation with a diverse set of personas to avoid drawing conclusions based on unrepresentative samples. Adhere to principles of fairness, diversity, and inclusivity in design and interpretation.[98]
- **Transparency and Explainability:** The complexity of LLMs and simulations can make them opaque ("black boxes"). Understanding how results are generated is crucial for trust and responsible interpretation.
    - *Mitigation:* Maintain meticulous documentation of the simulation logic, agent design principles, prompt structures, data sources, and parameters used.[98] The strategy of logging Agent 1's textual justifications [Insight 4.1] directly contributes to explainability by revealing the simulated reasoning process. Publish code and methods where possible (considering competition rules).
- **Potential for Misuse:** Findings from the simulation, particularly regarding effective influence strategies, could potentially be misinterpreted or misused for manipulative purposes (e.g., designing more effective disinformation campaigns).
    - *Mitigation:* Clearly articulate the simulation's limitations and the specific context of its findings in all reporting. Frame results responsibly, focusing on insights for

promoting public health and countering misinformation, rather than providing a "how-to" guide for manipulation. Emphasize the ethical responsibility associated with the research outcomes.[98] Consider the "dual-use" potential inherent in studying influence mechanisms.[98]

- **Data Privacy:** Even though the agents are synthetic, the data used to ground their personas (census data, survey results) must be handled ethically.
  - *Mitigation:* Prioritize the use of publicly available, aggregated, and anonymized data sources. If any potentially sensitive microdata is used for grounding, ensure it is fully anonymized and handled according to relevant privacy regulations and data use agreements. Ensure the simulation itself does not inadvertently leak any underlying sensitive data patterns. Adhere to principles of data agency and privacy.[98]
- **Adherence to Formal Guidelines:** Explicitly acknowledge and strive to adhere to relevant ethical codes and principles from professional organizations like the Association for Computing Machinery (ACM) [98] and the Institute of Electrical and Electronics Engineers (IEEE).[99] Key principles include professional responsibility, competence, honesty, avoiding harm, fairness, transparency, accountability, and respecting privacy. Reference ethical discussions specific to ABM in social science [98] and the emerging ethics of AI agent simulations, particularly in health contexts.[100]
- **The Duality of LLMs in Misinformation Research:** A specific ethical tension arises from using LLMs to *generate* potentially harmful content (misinformation) as part of the research process. LLMs are powerful tools that can be used to create persuasive misinformation [36], yet they are also being explored for detecting misinformation [36] and generating counter-messages or corrections.[31] This simulation places Agent 2 in the role of potentially generating misinformation to study its effects on Agent 1.
  This involves leveraging the technology's capability for problematic content generation within a controlled experimental setting.[36] The ethical justification rests heavily on the principle that understanding a phenomenon, even a harmful one, is often necessary to combat it effectively.[1] Ethical guidelines emphasize avoiding harm and considering misuse.[98] Therefore, the use of Agent 2 to generate misinformation must be strictly controlled, carefully justified by the research question, and aimed squarely at producing knowledge that can inform beneficial interventions (e.g., designing better public health communication or counter-misinformation strategies). The potential benefits of the research in promoting public health must clearly outweigh the risks associated with the methodology. Transparency about this approach and responsible reporting of the findings are paramount to navigate this ethical duality.

8. Conclusion and Recommendations for Competition

Summary of the Plan

This report has outlined a comprehensive plan for simulating the influence of targeted news on vaccine acceptance using a novel two-agent LLM framework. The core methodology involves:

- An **Agent-Based Modeling (ABM)** approach utilizing two interacting LLM agents: a 'Receiver' with a data-grounded persona and an 'Influencer' generating targeted news.
- **Rich Agent Design:** Leveraging real-world data (census, surveys, news corpora) and LLM synthesis to create detailed Receiver personas and guide the generation of realistic, targeted news content by the Influencer.
- **Iterative Simulation Loop:** Implementing a dynamic loop where the Influencer generates news based on the Receiver's current state, and the Receiver updates its belief (quantified on a Likert scale and explained via text justification) based on the news and its persona.
- **Technical Implementation:** Utilizing n8n on Azure for workflow orchestration, integrating with OpenAI and Google AI Studio APIs for LLM functionalities, and employing careful state management.
- **Multi-Method Evaluation:** Combining quantitative analysis of belief trajectories with qualitative analysis of agent reasoning, alongside checks for simulation validity and agent realism.
- **Ethical Awareness:** Proactively addressing ethical considerations related to bias, misinformation generation, transparency, and potential misuse.

Strengths for the Competition

This plan offers several advantages within a healthcare data analytics competition context:

- **Novelty and Technical Sophistication:** The use of interacting LLM agents to simulate belief dynamics represents a cutting-edge approach, blending AI, ABM, and computational social science. It moves beyond standard correlational analyses or simpler predictive models.
- **Methodological Rigor:** The plan emphasizes grounding the simulation in real-world data, employing established theoretical frameworks (ABM, opinion dynamics), and includes a robust multi-method evaluation strategy.
- **Technical Feasibility:** Provides a clear roadmap for implementation using the specific tools mandated by the user (n8n, OpenAI, Google AI Studio), addressing potential challenges like state management.
- **Direct Healthcare Relevance:** Tackles the critical public health issue of vaccine hesitancy and aims to provide actionable insights for designing more effective communication strategies, directly aligning with healthcare analytics goals.
- **Depth of Insight:** The ability to capture not just *if* beliefs change, but *why* (through qualitative analysis of agent reasoning), offers a deeper level of understanding than purely quantitative approaches.

Actionable Next Steps

To execute this plan, the following concrete steps are recommended:

1. **Prioritize Data Acquisition:** Begin immediately to identify, access, and process the required grounding data:

    - Secure public demographic data (e.g., Census).
    - Identify and obtain access to relevant public opinion surveys (e.g., Pew, KFF, GSS microdata if possible).
    - Gather seed examples of real/fake vaccine news (download datasets like CoVID19-FNIR [37], scrape ethically [78]).
    - Set up access to the Google Fact Check Tools API [90] if required for fact-checking integration.
2. **Develop Agent Prompts:** Draft detailed and structured prompts for:
    - Generating Agent 1 (Receiver) persona descriptions based on sampled parameters (Section 3.1).
    - Generating Agent 2 (Influencer) news snippets, incorporating targeting parameters and stylistic controls (Section 3.2).
    - Performing Agent 1's belief update, ensuring output includes both the numerical score and textual justification (Section 4.2).
3. **Build n8n Workflow:** Construct the core iterative loop in n8n as outlined in Section 4.3 and Table 2. Set up API connections to OpenAI/Google AI Studio and configure external state storage (recommended).
4. **Implement Core Logic:** Write the necessary functions (e.g., in n8n's Function/Code nodes) for parsing LLM responses, updating state, constructing dynamic prompts, and logging data.
5. **Conduct Pilot Runs:** Execute initial simulation runs with a small number of iterations and diverse test personas/strategies. Use these runs to debug the workflow, refine prompts for clarity and consistency, and test the state management mechanism.
6. **Execute Full Evaluation:** Once the simulation is stable, run experiments according to the evaluation plan (Section 6). Systematically vary persona parameters and news strategies. Collect and analyze both quantitative (belief scores) and qualitative (justifications) data.
7. **Document Thoroughly:** Maintain detailed records of methodology, data sources, prompt versions, code, simulation parameters, results, and ethical considerations throughout the process. This documentation is crucial for the competition submission and for reproducibility.

Framing for Competition Success

To maximize impact in the competition, frame the project by emphasizing:

- **The Innovative Methodology:** Highlight the novel application of LLM-driven generative agents within an ABM framework to tackle a complex socio-behavioral problem in healthcare.
- **Bridging AI and Public Health:** Position the work as a practical example of how advanced AI techniques can provide tangible insights for critical public health challenges like vaccine communication.
- **Beyond Prediction to Explanation:** Emphasize that the simulation aims not just to predict belief changes but to *explain* the underlying dynamics by analyzing the agents'

reasoning processes, offering deeper insights than traditional models.

- **Data-Driven Insights for Action:** Showcase how the simulation, grounded in real-world data, can generate specific, data-driven recommendations for tailoring communication strategies to different population segments, demonstrating clear potential for real-world application.
- **Technical Depth and Rigor:** Clearly present the technical architecture, the data grounding strategy, the multi-method evaluation plan, and the careful consideration of ethical issues to demonstrate a sophisticated and responsible approach to data analytics.

By effectively communicating these aspects, the project can demonstrate a powerful blend of technical innovation, methodological rigor, and direct relevance to improving healthcare outcomes through better communication.

Works cited

1. COVID-19 Vaccine Hesitancy and Information Diffusion: An Agent-based Modeling Approach - arXiv, accessed April 30, 2025, https://arxiv.org/pdf/2109.01182
2. Agent Based Model of Anti-Vaccination Movements: Simulations and Comparison with Empirical Data - PubMed Central, accessed April 30, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC8402338/
3. Can A Society of Generative Agents Simulate Human Behavior and Inform Public Health Policy? A Case Study on Vaccine Hesitancy - arXiv, accessed April 30, 2025, https://arxiv.org/html/2503.09639v1
4. arXiv:2503.09639v3 [cs.MA] 2 Apr 2025, accessed April 30, 2025, https://arxiv.org/pdf/2503.09639
5. Behavioral Dynamics of Epidemic Trajectories and Vaccination Strategies: A Simulation-Based Analysis - JASSS, accessed April 30, 2025, https://www.jasss.org/28/1/3/3.pdf
6. Simulating hidden dynamics | Amsterdam University Press Journals Online, accessed April 30, 2025, https://www.aup-online.com/content/journals/10.5117/CCR2020.1.001.WETT?crawler=true
7. AI Agents Simulate 1,052 Individuals' Personalities with Impressive Accuracy | Stanford HAI, accessed April 30, 2025, https://hai.stanford.edu/news/ai-agents-simulate-1052-individuals-personalities-impressive-accuracy
8. AgentSociety: Large-Scale Simulation of LLM-Driven Generative Agents Advances Understanding of Human Behaviors and Society - arXiv, accessed April 30, 2025, https://arxiv.org/html/2502.08691v1
9. Quantifying the Persona Effect in LLM Simulations - ResearchGate, accessed April 30, 2025, https://www.researchgate.net/publication/384205641_Quantifying_the_Persona_Effect_in_LLM_Simulations
10. Can A Society of Generative Agents Simulate Human Behavior and Inform Public

Health Policy? A Case Study on Vaccine Hesitancy - arXiv, accessed April 30, 2025, https://arxiv.org/html/2503.09639v2
11. Can A Society of Generative Agents Simulate Human Behavior and Inform Public Health Policy? — A Case Study on Vaccine Hesitancy - arXiv, accessed April 30, 2025, https://arxiv.org/html/2503.09639
12. Can A Society of Generative Agents Simulate Human Behavior and Inform Public Health Policy? — A Case Study on Vaccine Hesitancy - arXiv, accessed April 30, 2025, https://arxiv.org/html/2503.09639v3
13. Simulating Opinion Dynamics with Networks of LLM-based Agents - Agam Goyal, accessed April 30, 2025, https://agoyal0512.github.io/assets/pdf/2024.findings-naacl.211.pdf
14. LLM Agents | Prompt Engineering Guide, accessed April 30, 2025, https://www.promptingguide.ai/research/llm-agents
15. Do Large Language Models Solve the Problems of Agent-Based Modeling? A Critical Review of Generative Social Simulations - arXiv, accessed April 30, 2025, https://arxiv.org/html/2504.03274v1
16. On agent-based modeling and computational social science - Frontiers, accessed April 30, 2025, https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2014.00668/full
17. On agent-based modeling and computational social science - PMC, accessed April 30, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC4094840/
18. ACE: A Completely Agent-Based Modeling Approach (Tesfatsion), accessed April 30, 2025, https://faculty.sites.iastate.edu/tesfatsi/archive/tesfatsi/ace.htm
19. (PDF) On agent-based modeling and computational social science - ResearchGate, accessed April 30, 2025, https://www.researchgate.net/publication/264393841_On_agent-based_modeling_and_computational_social_science
20. Agent-Based Modeling Frameworks: Tools and Platforms for Complex Simulations, accessed April 30, 2025, https://smythos.com/ai-agents/agent-architectures/agent-based-modeling-frameworks/
21. LLM-AIDSim: LLM-Enhanced Agent-Based Influence Diffusion Simulation in Social Networks - MDPI, accessed April 30, 2025, https://www.mdpi.com/2079-8954/13/1/29
22. 10 Agent-Based Modeling Strategies for Boosting Simulation Accuracy - Number Analytics, accessed April 30, 2025, https://www.numberanalytics.com/blog/10-agent-based-modeling-strategies-boosting-simulation-accuracy
23. Mesa 3: Agent-based modeling with Python in 2025 - Open Journals, accessed April 30, 2025, https://www.theoj.org/joss-papers/joss.07668/10.21105.joss.07668.pdf
24. LLM agents: The ultimate guide 2025 | SuperAnnotate, accessed April 30, 2025, https://www.superannotate.com/blog/llm-agents
25. [2503.09639] Can A Society of Generative Agents Simulate Human Behavior and

Inform Public Health Policy? A Case Study on Vaccine Hesitancy - arXiv, accessed April 30, 2025, https://arxiv.org/abs/2503.09639
26. Dogmatism and Domination: A Simulation Study | Episteme | Cambridge Core, accessed April 30, 2025, https://www.cambridge.org/core/journals/episteme/article/dogmatism-and-domination-a-simulation-study/210EFE19A54901DAAD5542BDAD1A735C
27. Computational Models of Political Learning and Belief Polarization - D-Scholarship@Pitt, accessed April 30, 2025, https://d-scholarship.pitt.edu/46788/13/Dissertation_MinsuJang_ETDCombined_AfterRevision.pdf
28. Calibrating an Opinion Dynamics Model to Empirical Opinion Distributions and Transitions, accessed April 30, 2025, https://www.jasss.org/26/4/9.html
29. A theoretical framework for agent-based modelling of infectious ..., accessed April 30, 2025, https://www.researchgate.net/publication/390608572_A_theoretical_framework_for_agent-based_modelling_of_infectious_disease_dynamics_under_misinformation_and_vaccine_hesitancy
30. (PDF) Agent Based Model of Anti-Vaccination Movements: Simulations and Comparison with Empirical Data - ResearchGate, accessed April 30, 2025, https://www.researchgate.net/publication/353408041_Agent_Based_Model_of_Anti-Vaccination_Movements_Simulations_and_Comparison_with_Empirical_Data
31. Working with AI to persuade: Examining a large language model's ability to generate pro-vaccination messages - Stanford HCI Group, accessed April 30, 2025, https://hci.stanford.edu/publications/2023/Karinshak_CSCW23.pdf
32. Generative AI for vaccine misbelief correction: Insights from targeting extraversion and pseudoscientific beliefs - PubMed, accessed April 30, 2025, https://pubmed.ncbi.nlm.nih.gov/40086038/
33. Agent-Based Modeling of Vaccine Hesitancy: Exploring the Role of Trust, Policy, and Socioeconomic Factors | Request PDF - ResearchGate, accessed April 30, 2025, https://www.researchgate.net/publication/382702796_Agent-Based_Modeling_of_Vaccine_Hesitancy_Exploring_the_Role_of_Trust_Policy_and_Socioeconomic_Factors
34. Computational Analysis and Simulation of Empathic Behaviors: A Survey of Empathy Modeling with Behavioral Signal Processing Framework - PMC - PubMed Central, accessed April 30, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC5405768/
35. Self-Refine: Iterative Refinement with Self-Feedback for LLMs, accessed April 30, 2025, https://learnprompting.org/docs/advanced/self_criticism/self_refine
36. VaxGuard: A Multi-Generator, Multi-Type, and Multi-Role Dataset for Detecting LLM-Generated Vaccine Misinformation - ResearchGate, accessed April 30, 2025, https://www.researchgate.net/publication/389786311_VaxGuard_A_Multi-Generator_Multi-Type_and_Multi-Role_Dataset_for_Detecting_LLM-Generated_Vaccine_Misinformation

37. Covid-19 Fake News Infodemic Research Dataset (CoVID19-FNIR ..., accessed April 30, 2025, https://ieee-dataport.org/open-access/covid-19-fake-news-infodemic-research-dataset-covid19-fnir-dataset
38. A COVID-19 Rumor Dataset - PMC, accessed April 30, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC8200409/
39. COVID-19 Fake News Dataset - Papers With Code, accessed April 30, 2025, https://paperswithcode.com/dataset/covid-19-fake-news-dataset
40. How fact-checking works | Transparency Center, accessed April 30, 2025, https://transparency.meta.com/features/how-fact-checking-works
41. COVID-19 Fake News Dataset - Kaggle, accessed April 30, 2025, https://www.kaggle.com/datasets/arashnic/covid19-fake-news
42. Machine Learning Techniques for Fake News Detection - Preprints.org, accessed April 30, 2025, https://www.preprints.org/manuscript/202503.0325/v1
43. From Misinformation to Insight: Machine Learning Strategies for Fake News Detection, accessed April 30, 2025, https://www.mdpi.com/2078-2489/16/3/189
44. Fake News Detection: Comparative Evaluation of BERT-like Models and Large Language Models with Generative AI-Annotated Data - arXiv, accessed April 30, 2025, https://arxiv.org/html/2412.14276v1
45. Creating and Validating Synthetic Datasets for LLM Evaluation ..., accessed April 30, 2025, https://phoenix.arize.com/creating-and-validating-synthetic-datasets-for-llm-evaluation-experimentation/
46. Synthetic data: A secret ingredient for better language models - Red Hat, accessed April 30, 2025, https://www.redhat.com/en/blog/synthetic-data-secret-ingredient-better-language-models
47. [2503.14023] Synthetic Data Generation Using Large Language Models: Advances in Text and Code - arXiv, accessed April 30, 2025, https://arxiv.org/abs/2503.14023
48. Transferable text data distillation by trajectory matching - arXiv, accessed April 30, 2025, https://arxiv.org/html/2504.09818v2
49. Differentially Private Knowledge Distillation via Synthetic Text Generation - arXiv, accessed April 30, 2025, https://arxiv.org/abs/2403.00932
50. [2402.10379] DataDreamer: A Tool for Synthetic Data Generation and Reproducible LLM Workflows - arXiv, accessed April 30, 2025, https://arxiv.org/abs/2402.10379
51. [2410.18588] Knowledge Distillation Using Frontier Open-source LLMs: Generalizability and the Role of Synthetic Data - arXiv, accessed April 30, 2025, https://arxiv.org/abs/2410.18588
52. Distilled Self-Critique of LLMs with Synthetic Data: a Bayesian Perspective - arXiv, accessed April 30, 2025, https://arxiv.org/html/2312.01957v1
53. DataDreamer: A Tool for Synthetic Data Generation and Reproducible LLM Workflows, accessed April 30, 2025, https://arxiv.org/html/2402.10379v2
54. Scaling Laws of Synthetic Data for Language Models - arXiv, accessed April 30, 2025, https://arxiv.org/html/2503.19551

55. Approaches to Identify Fake News: A Systematic Literature Review - PMC, accessed April 30, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC7250114/
56. synthetic data generation using large language models: advances in text and code - arXiv, accessed April 30, 2025, https://arxiv.org/pdf/2503.14023
57. IMPROVE: Iterative Model Pipeline Refinement and Optimization Leveraging LLM Agents, accessed April 30, 2025, https://arxiv.org/html/2502.18530v1
58. RewardDS: Privacy-Preserving Fine-Tuning for Large Language Models via Reward Driven Data Synthesis - arXiv, accessed April 30, 2025, https://arxiv.org/html/2502.18517v1
59. Self-Boosting LLMs with Synthetic Preference Data - arXiv, accessed April 30, 2025, https://arxiv.org/html/2410.06961v1
60. LLM2LLM: Boosting LLMs with Novel Iterative Data Enhancement - arXiv, accessed April 30, 2025, https://arxiv.org/html/2403.15042v1
61. CodecLM: Aligning Language Models with Tailored Synthetic Data - arXiv, accessed April 30, 2025, https://arxiv.org/html/2404.05875v1
62. Iterative Deepening Sampling for Large Language Models - arXiv, accessed April 30, 2025, https://arxiv.org/html/2502.05449v1
63. A novel approach to fake news classification using LSTM-based deep learning models, accessed April 30, 2025, https://pmc.ncbi.nlm.nih.gov/articles/PMC10800750/
64. Covid-19 Fake News Detection: A Survey - SciELO México, accessed April 30, 2025, https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S1405-55462021000400783
65. Mastering LLM Prompting Techniques - DataRoot Labs, accessed April 30, 2025, https://datarootlabs.com/blog/prompting-techniques
66. Media and News Web Scraping | Use Cases and Methods - Scrapfly, accessed April 30, 2025, https://scrapfly.io/use-case/media-and-news-web-scraping
67. Mesa 3.0: A major update to Python's Agent-Based Modeling library - Reddit, accessed April 30, 2025, https://www.reddit.com/r/Python/comments/1gn5q8z/mesa_30_a_major_update_to_pythons_agentbased/
68. Choosing the Right LLM Agent Framework in 2025 - Botpress, accessed April 30, 2025, https://botpress.com/blog/llm-agent-framework
69. A curated list of awesome LLM agents frameworks. - GitHub, accessed April 30, 2025, https://github.com/kaushikb11/awesome-llm-agents
70. AutoGen — AutoGen, accessed April 30, 2025, https://microsoft.github.io/autogen/
71. Multi-agent LLMs in 2024 [+frameworks] | SuperAnnotate, accessed April 30, 2025, https://www.superannotate.com/blog/multi-agent-llms
72. CrewAI: Introduction, accessed April 30, 2025, https://docs.crewai.com/
73. CrewAI vs. AutoGen: Choosing the Right AI Agent Framework - Deepak Gupta, accessed April 30, 2025, https://guptadeepak.com/crewai-vs-autogen-choosing-the-right-ai-agent-framework/

74. OpenAI Agents SDK vs LangGraph vs Autogen vs CrewAI - Composio, accessed April 30, 2025, https://composio.dev/blog/openai-agents-sdk-vs-langgraph-vs-autogen-vs-crewai/
75. CrewAI vs. AutoGen: Comparing AI Agent Frameworks - Oxylabs, accessed April 30, 2025, https://oxylabs.io/blog/crewai-vs-autogen
76. CrewAI vs. AutoGen: Which Open-Source Framework is Better for Building AI Agents?, accessed April 30, 2025, https://www.helicone.ai/blog/crewai-vs-autogen
77. AutoGen vs. CrewAI: Compare AI agent frameworks - SmythOS, accessed April 30, 2025, https://smythos.com/ai-agents/comparison/autogen-vs-crewai/
78. 15 Python Web Scraping Projects: From Beginner to Advanced - Firecrawl, accessed April 30, 2025, https://www.firecrawl.dev/blog/python-web-scraping-projects
79. How to Scrape News Articles With Python and AI - Bright Data, accessed April 30, 2025, https://brightdata.com/blog/web-data/how-to-scrape-news-articles
80. Web Sites for Fact Checking - Misinformation and Disinformation: Thinking Critically about Information Sources, accessed April 30, 2025, https://library.csi.cuny.edu/c.php?g=619342&p=4310783
81. How to fact-check: free tools you can use to vet possible misinformation and improve media literacy - WebPurify, accessed April 30, 2025, https://www.webpurify.com/blog/how-to-fact-check-free-tools/
82. Using LLMs for Synthetic Data Generation: The Definitive Guide - Confident AI, accessed April 30, 2025, https://www.confident-ai.com/blog/the-definitive-guide-to-synthetic-data-generation-using-llms
83. Snopes - RAND, accessed April 30, 2025, https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search/items/snopes.html
84. Fact-checking - Wikipedia, accessed April 30, 2025, https://en.wikipedia.org/wiki/Fact-checking
85. No more fact-checking for Meta. How will this change media — and the pursuit of truth?, accessed April 30, 2025, https://www.ap.org/news-highlights/spotlights/2025/no-more-fact-checking-for-meta-how-will-this-change-media-and-the-pursuit-of-truth/
86. The perils and promises of fact-checking with large language models - Frontiers, accessed April 30, 2025, https://www.frontiersin.org/journals/artificial-intelligence/articles/10.3389/frai.2024.1341697/full
87. News Verified and Fact-Checked with Reuters, accessed April 30, 2025, https://reutersagency.com/solutions/verification-services/
88. About Us - Reuters, accessed April 30, 2025, https://reutersagency.com/about/
89. Reuters launches fact-checking initiative to identify misinformation, in partnership with Facebook, accessed April 30, 2025, https://www.thomsonreuters.com/en/press-releases/2020/february/reuters-launc

hes-fact-checking-initiative-to-identify-misinformation
90. Fact Check Tools API | Google for Developers, accessed April 30, 2025, https://developers.google.com/fact-check/tools/api
91. Package google.factchecking.factchecktools.v1alpha1 | Fact Check Tools API, accessed April 30, 2025, https://developers.google.com/fact-check/tools/api/reference/rpc/google.factchecking.factchecktools.v1alpha1
92. Google Fact Check Tool APIs, accessed April 30, 2025, https://toolbox.google.com/factcheck/apis?hl=ar
93. Python Web Scraping: Full Tutorial With Examples (2025) - ScrapingBee, accessed April 30, 2025, https://www.scrapingbee.com/blog/web-scraping-101-with-python/
94. Synthetic Data Generator Simplifies Dataset Creation with Large Language Models - InfoQ, accessed April 30, 2025, https://www.infoq.com/news/2025/01/synthetic-data-generator/
95. Mesa: Agent-based modeling in Python — Mesa .1 documentation, accessed April 30, 2025, https://mesa.readthedocs.io/en/stable/
96. Ensuring Accuracy and Equity in Vaccination Information From ChatGPT and CDC: Mixed-Methods Cross-Language Evaluation, accessed April 30, 2025, https://formative.jmir.org/2024/1/e60939
97. Iterative Prompt Refinement: Step-by-Step Guide - Ghost, accessed April 30, 2025, https://latitude-blog.ghost.io/blog/iterative-prompt-refinement-step-by-step-guide/
98. The Ethics of Agent-Based Social Simulation - JASSS, accessed April 30, 2025, https://www.jasss.org/25/4/1.html
99. sagroups.ieee.org, accessed April 30, 2025, https://sagroups.ieee.org/global-initiative/wp-content/uploads/sites/542/2023/01/ead1e.pdf
100. (PDF) Personas Evolved: Designing Ethical LLM-Based Conversational Agent Personalities, accessed April 30, 2025, https://www.researchgate.net/publication/389510663_Personas_Evolved_Designing_Ethical_LLM-Based_Conversational_Agent_Personalities
101. Information Security, Ethics, and Integrity in LLM Agent Interaction, accessed April 30, 2025, https://www.scirp.org/journal/paperinformation?paperid=140224
102. A Survey on LLM-based Multi-Agent AI Hospital - OSF, accessed April 30, 2025, https://osf.io/bv5sg_v1/download/?format=pdf