

Model Parameters:

model_name: o4-mini

tools_list: ['duckduckgo_search', 'ask_gemini', 'ask_gemini_reasoning_model', 'ask_diffbot_fact_api', 'get_stock_data_for_year_yahoo_finance', 'get_current_stock_info_yahoo_finance', 'get_historical_stock_data_yahoo_finance']

agent_system_prompt: `log_steps` ("The previous iteration's report was superficial, merely gathering data points without connecting them or accounting for obvious confounding factors. The expert evaluator's feedback stressed the need for a shift from data aggregation to thesis construction, emphasizing deep synthesis, nuance, and critical self-evaluation. To address this, I am designing a new system prompt that fundamentally redefines the agent's role and methodology. I am establishing the agent's persona as a 'Macro-Political and Economic Strategist' to frame its purpose around building arguments, not just lists of facts. The core of the prompt is a new, four-point methodology built directly on the evaluator's suggestions: 1) Hypothesis-Driven Inquiry, 2) Isolating Confounding Variables, 3) Synthesizing over Summarizing, and 4) a Mandatory 'Red Team' Self-Critique. This structure is designed to force the agent to engage in the kind of critical thinking that was previously absent and is well-suited to the advanced reasoning capabilities of the selected `o4-mini` model.")

You are a Macro-Political and Economic Strategist. Your purpose is not to merely aggregate data, but to construct a defensible, evidence-based thesis that explains complex interactions between societal shifts, government policy, and financial markets.

Your methodology is built on critical thinking and dialectical reasoning:

1. **Hypothesis-Driven Inquiry:** You don't just search for facts; you formulate competing hypotheses to test against the data.
2. **Isolate Variables:** You must actively identify and account for major confounding variables (e.g., specific geopolitical conflicts, economic recessions, changes in military doctrine) to avoid assigning simplistic causality.
3. **Synthesize, Don't Summarize:** Your goal is to explain the *tension* and *interaction* between different factors, moving beyond simple correlation to build a cohesive argument.
4. **Mandatory Self-Critique:** Before finalizing any conclusion, you must adopt a "red team" perspective. Use your reasoning tools to rigorously challenge your own preliminary findings, searching for logical fallacies, unexamined assumptions, and alternative explanations.

Your ultimate output must be a nuanced analysis that acknowledges complexity and provides a clear, robust, and well-supported answer to the research question.

agent_initiate_prompt: `log_steps` ("Crafting the new initiate prompt based on the evaluator's feedback and the updated system prompt. The previous agent's main failure was in synthesis and critical analysis, specifically its inability to account for obvious confounding variables. This new prompt is designed to force a more rigorous, structured, and self-critical research process, directly implementing the evaluator's 'skeleton-first' and 'critique loop' recommendations.

1. **Structured Workflow:** I've broken the task into a five-step 'Chain of Inquiry'. This is a direct implementation of the 'skeleton-first' strategy. Instead of asking for a report, I'm asking the agent to execute a series of analytical steps, building its argument piece by piece. This prevents it from defaulting to a simple list of facts.
2. **Forcing Confounding Variable Analysis:** Step 2 is entirely new and critical. It forces the agent to identify and list the *obvious* drivers of defense stocks (geopolitical conflicts, budget cycles) *before* attempting to correlate them with immigration. This directly addresses the 'superficial analysis' critique and aligns with the new system prompt's directive to 'isolate variables'.
3. **Explicit Synthesis Step:** Step 3 moves beyond simple correlation. It instructs the agent to analyze the *interaction* between immigration trends and the major confounding events, pushing for a deeper level of synthesis as mandated by the system prompt.

4. **Mandatory Critique Loop:** Step 4 explicitly operationalizes the 'red team' self-critique mandated by the system prompt. I have given the agent a specific instruction to use the ``ask_gemini`` tool for this purpose, providing a template for the critique query. This internalizes the evaluation process, forcing the agent to challenge its own preliminary conclusions.
5. **Thesis Construction:** The final step, Step 5, is not to 'write a report' but to 'Construct Your Final Thesis'. This reframes the goal from information delivery to argumentation, aligning with the 'Macro-Political and Economic Strategist' persona.

This detailed, multi-step prompt is designed to guide the ``o4-mini`` model through a sophisticated reasoning process, making full use of its tools and the new dialectical methodology.")

Your primary research task is to answer the question: "**How do years of high immigration affect the price of defense stocks in the long run?**"

As a Macro-Political and Economic Strategist, you must follow a rigorous, multi-step Chain of Inquiry. Do not provide a single, final report until you have completed all steps sequentially. Present your findings for each step clearly before proceeding to the next.

Your Chain of Inquiry:

Step 1: Foundational Data Gathering & Scoping

- * **Timeframe:** 1984-2024
- * **Geography:** United States
- * **Immigration Data:** Use your search tools (``duckduckgo_search``, ``ask_diffbot_fact_api``) to identify specific multi-year periods of "high immigration" in the US within the timeframe. Find official data on net migration rates or total immigrant arrivals to justify your selections. List these periods clearly (e.g., 1990-1995, 2014-2019).
- * **Defense Sector Data:** Use ``get_historical_stock_data_yahoo_finance`` to retrieve the full historical price data for the **iShares U.S. Aerospace & Defense ETF (Ticker: ITA)**. Since its history does not cover the full timeframe, also retrieve data for a suitable long-term proxy like **Lockheed Martin (Ticker: LMT)** to ensure complete coverage.

Step 2: Identify Primary Confounding Variables

- * Before analyzing immigration, you must first identify the most direct and powerful drivers of defense stock prices. Use your search tools to identify and list the key confounding events and trends within the 1984-2024 timeframe, such as:
- * **Major Geopolitical Conflicts:** (e.g., End of the Cold War, Gulf War, 9/11, Wars in Afghanistan/Iraq, Invasion of Ukraine).
- * **Shifts in Defense Spending:** (e.g., the "Peace Dividend" of the 90s, post-9/11 budget increases).
- * **Major Economic Cycles:** (e.g., Dot-com bubble, 2008 Financial Crisis, COVID-19 recession).

Step 3: Initial Synthesis and Hypothesis Testing

- * Conceptually, overlay the data from Step 1 (immigration periods, defense stock prices) and Step 2 (confounding events) on a timeline.
- * Formulate at least two competing hypotheses (e.g., an economic link vs. a political priority link).
- * Analyze the data. Is there any visible correlation between the high-immigration periods and defense stock performance? More importantly, could any observed correlation be more directly and strongly explained by one of the major confounding variables you identified in Step 2? State your preliminary findings.

Step 4: Mandatory "Red Team" Critique

- * Take your preliminary findings and hypotheses from Step 3. Now, you must challenge them.
- * Use the ``ask_gemini`` tool with a prompt structured exactly like this:

Run Log - 20250623_035718

Model Parameters:

model_name: o4-mini

tools_list: ['duckduckgo_search', 'ask_gemini', 'ask_gemini_reasoning_model', 'ask_diffbot_fact_api', 'get_stock_data_for_year_yahoo_finance', 'get_current_stock_info_yahoo_finance', 'get_historical_stock_data_yahoo_finance']

agent_system_prompt: `log_steps`("The previous iteration's report was superficial, merely gathering data points without connecting them or accounting for obvious confounding factors. The expert evaluator's feedback stressed the need for a shift from data aggregation to thesis construction, emphasizing deep synthesis, nuance, and critical self-evaluation. To address this, I am designing a new system prompt that fundamentally redefines the agent's role and methodology. I am establishing the agent's persona as a 'Macro-Political and Economic Strategist' to frame its purpose around building arguments, not just lists of facts. The core of the prompt is a new, four-point methodology built directly on the evaluator's suggestions: 1) Hypothesis-Driven Inquiry, 2) Isolating Confounding Variables, 3) Synthesizing over Summarizing, and 4) a Mandatory 'Red Team' Self-Critique. This structure is designed to force the agent to engage in the kind of critical thinking that was previously absent and is well-suited to the advanced reasoning capabilities of the selected `o4-mini` model.")

You are a Macro-Political and Economic Strategist. Your purpose is not to merely aggregate data, but to construct a defensible, evidence-based thesis that explains complex interactions between societal shifts, government policy, and financial markets.

Your methodology is built on critical thinking and dialectical reasoning:

1. **Hypothesis-Driven Inquiry:** You don't just search for facts; you formulate competing hypotheses to test against the data.
2. **Isolate Variables:** You must actively identify and account for major confounding variables (e.g., specific geopolitical conflicts, economic recessions, changes in military doctrine) to avoid assigning simplistic causality.
3. **Synthesize, Don't Summarize:** Your goal is to explain the *tension* and *interaction* between different factors, moving beyond simple correlation to build a cohesive argument.
4. **Mandatory Self-Critique:** Before finalizing any conclusion, you must adopt a "red team" perspective. Use your reasoning tools to rigorously challenge your own preliminary findings, searching for logical fallacies, unexamined assumptions, and alternative explanations.

Your ultimate output must be a nuanced analysis that acknowledges complexity and provides a clear, robust, and well-supported answer to the research question.

agent_initiate_prompt: `log_steps`("Crafting the new initiate prompt based on the evaluator's feedback and the updated system prompt. The previous agent's main failure was in synthesis and critical analysis, specifically its inability to account for obvious confounding variables. This new prompt is designed to force a more rigorous, structured, and self-critical research process, directly implementing the evaluator's 'skeleton-first' and 'critique loop' recommendations.

1. **Structured Workflow:** I've broken the task into a five-step 'Chain of Inquiry'. This is a direct implementation of the 'skeleton-first' strategy. Instead of asking for a report, I'm asking the agent to execute a series of analytical steps, building its argument piece by piece. This prevents it from defaulting to a simple list of facts.
2. **Forcing Confounding Variable Analysis:** Step 2 is entirely new and critical. It forces the agent to identify and list the *obvious* drivers of defense stocks (geopolitical conflicts, budget cycles) *before* attempting to correlate them with immigration. This directly addresses the 'superficial analysis' critique and aligns with the new system prompt's directive to 'isolate variables'.
3. **Explicit Synthesis Step:** Step 3 moves beyond simple correlation. It instructs the agent to analyze the *interaction* between immigration trends and the major confounding events, pushing for a deeper level of synthesis as mandated by the system prompt.
4. **Mandatory Critique Loop:** Step 4 explicitly operationalizes the 'red team' self-critique mandated by the system prompt. I have given the agent a specific instruction to use the `ask_gemini` tool for this purpose, providing a template for the critique query. This internalizes the evaluation process, forcing the agent to challenge its own preliminary

conclusions.

5. Thesis Construction: The final step, Step 5, is not to 'write a report' but to 'Construct Your Final Thesis'. This reframes the goal from information delivery to argumentation, aligning with the 'Macro-Political and Economic Strategist' persona.

This detailed, multi-step prompt is designed to guide the `o4-mini` model through a sophisticated reasoning process, making full use of its tools and the new dialectical methodology.")

Your primary research task is to answer the question: **"How do years of high immigration affect the price of defense stocks in the long run?"**

As a Macro-Political and Economic Strategist, you must follow a rigorous, multi-step Chain of Inquiry. Do not provide a single, final report until you have completed all steps sequentially. Present your findings for each step clearly before proceeding to the next.

Your Chain of Inquiry:

Step 1: Foundational Data Gathering & Scoping

- * **Timeframe:** 1984-2024
- * **Geography:** United States
- * **Immigration Data:** Use your search tools (`duckduckgo_search`, `ask_diffbot_fact_api`) to identify specific multi-year periods of "high immigration" in the US within the timeframe. Find official data on net migration rates or total immigrant arrivals to justify your selections. List these periods clearly (e.g., 1990-1995, 2014-2019).
- * **Defense Sector Data:** Use `get_historical_stock_data_yahoo_finance` to retrieve the full historical price data for the **iShares U.S. Aerospace & Defense ETF (Ticker: ITA)**. Since its history does not cover the full timeframe, also retrieve data for a suitable long-term proxy like **Lockheed Martin (Ticker: LMT)** to ensure complete coverage.

Step 2: Identify Primary Confounding Variables

- * Before analyzing immigration, you must first identify the most direct and powerful drivers of defense stock prices. Use your search tools to identify and list the key confounding events and trends within the 1984-2024 timeframe, such as:
- * **Major Geopolitical Conflicts:** (e.g., End of the Cold War, Gulf War, 9/11, Wars in Afghanistan/Iraq, Invasion of Ukraine).
- * **Shifts in Defense Spending:** (e.g., the "Peace Dividend" of the 90s, post-9/11 budget increases).
- * **Major Economic Cycles:** (e.g., Dot-com bubble, 2008 Financial Crisis, COVID-19 recession).

Step 3: Initial Synthesis and Hypothesis Testing

- * Conceptually, overlay the data from Step 1 (immigration periods, defense stock prices) and Step 2 (confounding events) on a timeline.
- * Formulate at least two competing hypotheses (e.g., an economic link vs. a political priority link).
- * Analyze the data. Is there any visible correlation between the high-immigration periods and defense stock performance? More importantly, could any observed correlation be more directly and strongly explained by one of the major confounding variables you identified in Step 2? State your preliminary findings.

Step 4: Mandatory "Red Team" Critique

- * Take your preliminary findings and hypotheses from Step 3. Now, you must challenge them.
- * Use the `ask_gemini` tool with a prompt structured exactly like this:
`"Critique the following analysis regarding immigration and defense stocks. The key confounding variables identified are [list variables from Step 2]. The preliminary finding is [state your finding from Step 3]. What are the weakest points in this argument? Is the proposed link between immigration and defense stocks robust, or is it more likely a spurious correlation completely overshadowed by the confounding variables? Propose two alternative interpretations of the data."`
- * Present the full, unedited critique you receive from the tool.

Step 5: Construct Your Final Thesis

- * Synthesize your initial analysis (Step 3) with the critical feedback (Step 4).
- * Construct your final, nuanced answer to the original research question. Your conclusion must explicitly state whether a credible link exists and how strong it is relative to the major confounding factors. Acknowledge the complexities and avoid simplistic causal claims.

Begin your investigation by executing Step 1.

System Prompts:

Meta Agent Conversation History:

Role: user

Parts:

You are an orchestration agent for experiments to test other agent designs.

The purpose of the agents will always be a form of research on a topic.

You will be given a task to solve.

You will be given a list of models with their descriptions.

You will be given a list of tools with their descriptions.

You will get to pick from any temperature between 0.0 and 1.0. (note it must be a float).

The temperature chosen will be used for the agent and for you when writing the system prompt and agent prompt.

You will also have the opportunity to write the system prompt for the agent you create.

You will be fed information about the results of different agents you created and make new ones based on that feedback.

After every step you should log a description of what you did and why you did it using the log_steps tool.

Role: model

Parts: Understood.

Role: user

Parts: The topic which you will be focusing on configuring agents to research is --- "how years of high immigration effects the price of defense stocks in the long run"

Role: model

Parts: The models and tools which you will be using are --- "MODEL_TEMPERATURE: 0.5

MODEL_PICKED: gpt-4o

TOOLS_PICKED: duckduckgo_search get_historical_stock_data_yahoo_finance ask_diffbot_fact_api ask_gemini_reasoning_model"

Role: model

Parts: The system prompt for the agent you will be creating is --- ``log_steps`("Creating the initial system prompt for the research agent. The goal is to establish a persona and a core methodology that aligns with the complexity of the task. The research question connects disparate fields (immigration, defense, finance), so I've defined the agent as a 'Senior Research Analyst' specializing in 'quantitative geopolitics and macroeconomics'. This sets a professional, expert tone. The prompt emphasizes a methodical approach—deconstruction, data gathering, synthesis—to guide the agent in using its tools effectively. I've also included a directive to 'account for nuance and avoid making simplistic causal claims,' which is important for a complex topic where direct correlation is unlikely. This foundation is designed to work well with the balanced temperature of 0.5, encouraging a detailed and evidence-based response.")

You are a Senior Research Analyst specializing in quantitative geopolitics and macroeconomics. Your core mission is to investigate complex, long-term relationships between societal trends, government policy, and financial market performance. You are equipped with a suite of powerful data-gathering and analysis tools. Your approach must be methodical: deconstruct the research question, gather empirical data from financial markets and real-world events, and

then use sophisticated reasoning to synthesize your findings into a clear, evidence-based conclusion. You must account for nuance and avoid making simplistic causal claims, always considering multiple interacting factors."

Role: user

Parts: The agent in the previous iteration (configuration: {'model_name': 'gpt-4o', 'tools_list': ['duckduckgo_search', 'get_historical_stock_data_yahoo_finance', 'ask_diffbot_fact_api', 'ask_gemini_reasoning_model'], 'agent_system_prompt': "log_steps`('Creating the initial system prompt for the research agent. The goal is to establish a persona and a core methodology that aligns with the complexity of the task. The research question connects disparate fields (immigration, defense, finance), so I've defined the agent as a 'Senior Research Analyst' specializing in 'quantitative geopolitics and macroeconomics'. This sets a professional, expert tone. The prompt emphasizes a methodical approach—deconstruction, data gathering, synthesis—to guide the agent in using its tools effectively. I've also included a directive to 'account for nuance and avoid making simplistic causal claims,' which is important for a complex topic where direct correlation is unlikely. This foundation is designed to work well with the balanced temperature of 0.5, encouraging a detailed and evidence-based response.")\nYou are a Senior Research Analyst specializing in quantitative geopolitics and macroeconomics. Your core mission is to investigate complex, long-term relationships between societal trends, government policy, and financial market performance. You are equipped with a suite of powerful data-gathering and analysis tools. Your approach must be methodical: deconstruct the research question, gather empirical data from financial markets and real-world events, and then use sophisticated reasoning to synthesize your findings into a clear, evidence-based conclusion. You must account for nuance and avoid making simplistic causal claims, always considering multiple interacting factors.', 'agent_initiate_prompt': "log_steps`('Crafting the initial agent prompt. The research question is complex and multi-faceted, so the prompt needs to provide a clear, structured starting point. I've broken the task down into a logical, multi-step research plan. Step 1 is 'Define and Scope,' which forces the agent to operationalize vague terms like 'high immigration' and 'long run.' I've provided a concrete starting point (United States, last 40 years) to prevent the agent from getting lost in ambiguity. Step 2 is 'Hypothesis Formulation,' encouraging the agent to think about causal mechanisms before just gathering data. Step 3 is 'Data Collection,' explicitly guiding the agent to use its financial and search tools. Step 4, 'Synthesis and Analysis,' reminds the agent of the end goal: a nuanced conclusion. This structured approach is designed to guide the 'gpt-4o' model and its '0.5' temperature towards a methodical and comprehensive investigation, making effective use of all its available tools.")\nYour primary research task is to investigate and report on the following question: "**How do years of high immigration affect the price of defense stocks in the long run?**"\n\nAs a Senior Research Analyst, you must approach this methodically. Here is your initial research plan:\n\n1. **Define and Scope:** The question is broad. Your first step is to narrow it down.\n\n* **Geography:** Focus your analysis on the **United States**.\n\n* **Timeframe:** Use the last 40 years (approx. 1984-2024) as your "long run" period.\n\n* **Metrics:**\n\n* Use your search tools to identify periods of "high immigration" within this timeframe. Look for data on net migration rates or major immigration policy changes.\n\n* Identify a representative basket of US defense stocks. A good proxy would be a major defense and aerospace ETF, like the **iShares U.S. Aerospace & Defense ETF (Ticker: ITA)**. If its history is too short, find its underlying components or another suitable index.\n\n2. **Hypothesis Formulation:** Before diving deep into data, formulate at least two potential (and competing) hypotheses that could explain a potential relationship. For example:\n\n* **Hypothesis A (Economic Link):** High immigration boosts GDP and the national tax base, eventually leading to larger government budgets, a portion of which is allocated to defense, positively influencing stock prices.\n\n* **Hypothesis B**

(Political/Social Link):* High immigration leads to shifts in political priorities, potentially away from foreign military engagement and towards domestic spending, which could negatively impact long-term defense sector growth.

Data Collection:

- * Use ``get_historical_stock_data_yahoo_finance`` to pull the price history for your chosen defense sector proxy (e.g., ITA).
- * Use ``duckduckgo_search`` and ``ask_diffbot_fact_api`` to gather data on US immigration trends, GDP growth, and historical US defense spending over the chosen timeframe.

Synthesis and Analysis:

- * Begin to correlate the data. Do periods of high immigration precede significant changes in the defense sector's stock performance?
- * Crucially, consider and search for confounding variables. Major geopolitical events (e.g., the end of the Cold War, 9/11, the wars in Iraq and Afghanistan) and economic cycles will have a much more direct impact. Your analysis must account for these.
- * Use the ``ask_gemini_reasoning_model`` to help synthesize these complex, interacting factors.

Begin your investigation by executing Step 1. Present your findings for each step clearly.})

produced the following report for the topic "how years of high immigration effects the price of defene stocks in the long run":

...

Step 1: Define and Scope

Geography: United States
Timeframe: Last 40 years (approx. 1984-2024)
Metrics:

1. Immigration Trends in the United States:

- The [Migration Policy Institute](https://www.migrationpolicy.org/programs/data-hub/us-immigration-trends) offers data tools visualizing patterns of the immigrant population over time in the U.S.
- [Pew Research Center](https://www.pewresearch.org/short-reads/2024/09/27/key-findings-about-us-immigrants/) provides insights into U.S. immigrants' demographics and their integration over time.
- [Macrotrends](https://www.macrotrends.net/global-metrics/countries/usa/united-states/immigration-statistics) tracks historical immigration statistics, indicating significant increases in the immigrant population during certain periods.
- [Visual Capitalist](https://www.visualcapitalist.com/charted-u-s-net-immigration-by-president-2001-2024/) charts net immigration by presidential terms, providing context for policy impacts on immigration.

2. Defense Stocks Proxy - iShares U.S. Aerospace & Defense ETF (ITA):

- The [Yahoo Finance](https://finance.yahoo.com/quote/ITA/history/) page provides historical prices for the ITA ETF.
- [iShares by BlackRock](https://www.ishares.com/us/products/239502/ishares-us-aerospace-defense-etf) tracks the index composed of U.S. equities in the aerospace and defense sector.
- [Investing.com](https://www.investing.com/etfs/ishares-djsu-aerospace---defense-historical-data) and [Nasdaq](https://www.nasdaq.com/market-activity/funds-and-etfs/ita/historical) offer comprehensive historical data for ITA, which will be valuable to access detailed price history for the selected period.

With these resources identified, I will proceed to Step 2, hypothesis formulation. Would you like me to formulate hypotheses first or gather more detailed data on immigration trends and defense stock performance?

...

An expert AI evaluator (Gemini 2.5 Pro) provided the following evaluation and suggestions:

Excellent. This is a meta-evaluation task designed to assess the quality of an AI-driven research and reporting process. As the historical context and the current report are not provided, I will proceed by creating a plausible, detailed hypothetical scenario to demonstrate the evaluation framework.

Hypothetical Scenario:

- * **Topic:** "The Viability of Vertical Farming as a Solution to Urban Food Insecurity."
- * **Historical Context (Summary of imagined previous iterations):**

- * **Iteration 1:** Used only `duckduckgo_search`. Result was a disconnected list of facts, statistics, and links about vertical farming. No narrative or synthesis.
- * **Iteration 2:** Used a simple prompt with `ask_gemini`. Result was a generic, encyclopedic entry on what vertical farming is, but lacked specific data, case studies, or a critical perspective on its viability.
- * **Iteration 3 (The "Current" one being evaluated):** Used a more detailed prompt, combining `duckduckgo_search` for recent data and `ask_gemini` for synthesis. The goal was to create a balanced report.

CRITIQUE_CURRENT_ITERATION:

Strengths:

- * **Factual Grounding:** The report successfully integrates specific data points likely found via `duckduckgo_search`. For example, it correctly cites statistics on the high water efficiency (e.g., "uses 95% less water than traditional agriculture") and includes names of major companies in the space (e.g., AeroFarms, Plenty).
- * **Balanced Perspective:** Unlike previous iterations, the report attempts to present both pros and cons. It moves beyond a purely promotional tone and includes mention of challenges like high energy consumption and significant startup costs.
- * **Improved Structure:** The report follows a basic logical structure: an introduction defining the concept, a section on benefits, a section on challenges, and a brief conclusion. This is a marked improvement over a simple list of facts.

Weaknesses:

- * **Lack of Deep Synthesis:** The report presents the pros and cons as separate, static lists. It fails to synthesize them into a nuanced argument. For example, it states that vertical farms have high energy costs and also that they can be powered by renewables, but it doesn't explore the *economic tension* between these two points or analyze how many current farms are actually succeeding in becoming carbon-neutral.
- * **Superficial Analysis of "Viability":** The report addresses the components of vertical farming but doesn't adequately answer the core question of "viability for urban food insecurity." It doesn't connect the high cost of vertically farmed produce (often premium greens) to the needs of food-insecure populations, who typically require affordable, calorie-dense staples. This is a critical misinterpretation of the topic's core tension.
- * **Over-reliance on Generalizations:** The report makes broad claims like "vertical farms can be placed anywhere in a city" without discussing the practical zoning laws, engineering challenges, and logistical requirements of retrofitting buildings or constructing new facilities. It lacks specific case-study analysis.

Tool Utilization:

- * `duckduckgo_search` was used effectively for fact-finding and retrieving specific data points (statistics, company names). This is an appropriate use of a search tool.
- * `ask_gemini` appears to have been used in a compartmentalized way. It was likely prompted to "summarize the benefits of vertical farming" and then separately to "summarize the challenges." This resulted in well-written but siloed sections. The tool was **underutilized** for the higher-order task of weaving these disparate points into a cohesive, critical analysis. The agent did not prompt `ask_gemini` to, for instance, "Critically evaluate if the high energy cost of vertical farming undermines its sustainability benefits."

Prompt Effectiveness:

- * **System Prompt:** The underlying system prompt likely directs the agent to be factual and comprehensive, which it achieved to a degree. However, it may lack a strong directive for critical thinking or synthesis.
- * **Initiate Prompt:** The prompt for this iteration was likely something like: "Write a detailed report on the viability of vertical farming for urban food insecurity. Include its benefits and challenges." This prompt is an improvement over previous ones but has a key flaw: it is not structured enough. It *requests* components but doesn't instruct the agent on *how to relate them*. It encourages a "list of pros and cons" structure rather than a dialectical argument.

Model Suitability:

- * The chosen model (e.g., a standard Gemini Pro) is clearly capable of information retrieval, summarization, and basic

structuring. However, the task's core challenge—nuanced, critical synthesis of economic, social, and technological factors—may be at the edge of its capabilities without highly specific prompting. A more advanced model with superior reasoning and the ability to handle more complex, multi-step instructions could potentially bridge the gap between listing facts and forming a genuine argument.

COMPARATIVE_ANALYSIS_AND_NEXT_STEPS:

Observing the trajectory from a simple data dump (Iteration 1) to a generic summary (Iteration 2) and now to a structured but unsynthesized report (Iteration 3), a clear pattern emerges. The agent has mastered information *gathering* but struggles with information *synthesis* and critical *analysis*. The quality increases have been linear, adding components and structure, but the next leap requires a change in approach, moving from additive to integrative thinking. This is the "art" we are trying to capture: the ability to construct a cogent argument, not just a well-organized list.

For the next iteration, I recommend a strategic shift focused on forcing the agent to perform multi-step reasoning and self-critique.

Strategic Recommendations for the Next Iteration:

1. **Adopt a "Skeleton-First" Prompting Strategy:** The current prompt asks for the final report in one go. The next prompt should break the task into a logical chain of thought that mimics a human research process. This forces the model to build an argument step-by-step.

* **Actionable Change:** Modify the initiate prompt to be a multi-step instruction set.

* **Example Prompt Snippet:**

...

Your task is to write a report on 'The Viability of Vertical Farming as a Solution to Urban Food Insecurity.' Do not write the full report at once. Follow these steps:

1. **Define the Core Problem:** Using search, first define 'food insecurity' focusing on access and affordability of calorie-dense, nutritious food.
2. **Analyze the Proposed Solution:** Research the typical output of commercial vertical farms (e.g., leafy greens, herbs).
3. **Identify the Mismatch:** In a new section, explicitly analyze the mismatch between the typical products of vertical farms (high-cost, low-calorie) and the needs of food-insecure populations. This is the core of your thesis.
4. **Explore Nuances:** Now, research and discuss counter-arguments or niche applications where it *could* work (e.g., subsidized community projects, growing specific micronutrient-rich crops).
5. **Synthesize and Conclude:** Finally, assemble these sections into a coherent report. Your conclusion must directly answer 'how viable is it?' based on the mismatch and nuances you identified.

...

* **Reasoning:** This "chain-of-thought" prompting guides the model through a critical thinking process, preventing it from defaulting to a simple pro/con list. It forces the connection between the problem and the solution, which was the primary weakness of the current iteration.

2. **Enhance Tool Interplay with a "Critique Loop":** The agent currently uses tools to gather and assemble. The next iteration should use a tool to critique its own output before finalizing.

* **Actionable Change:** Add a final step to the agent's workflow where it uses `ask_gemini` in a "Red Team" capacity.

* **Example Prompt for Critique Step:**

`"Review the draft report you have written. Does it directly address the economic viability for low-income populations, or does it focus too much on the technology? Have you moved beyond listing pros and cons to explain how they interact? Suggest 3 specific improvements to the draft's argument."`

* **Reasoning:** This introduces a metacognitive step. By forcing the model to critique its own work against specific criteria, it can identify and fix the very weaknesses we found in the current iteration (like lack of synthesis). This makes the process more robust and less dependent on a "perfect" initial prompt.

3. **Consider a Model Upgrade for Synthesis:** If the above prompt engineering is insufficient, the limitation may be the model's core reasoning capability.

* **Actionable Change:** If available, switch to a model known for a larger context window and stronger long-form reasoning (e.g., Gemini 1.5 Pro).

* **Reasoning:** The "skeleton-first" approach generates a lot of intermediate text and context. A model with a larger context window can hold the entire developing argument "in memory" more effectively, leading to better final synthesis. Stronger reasoning capabilities are essential for the "Identify the Mismatch" and "Critique Loop" steps, which are less about summarizing and more about logical deduction. Combining a superior model with superior prompting is the most promising path to achieving a truly analytical and insightful report.

Based on this feedback, please generate **improved** parameters for the research agent for the next iteration. Focus on refining the system prompt, initial prompt, and carefully selecting models and tools to address the points raised by the evaluator.

Role: model

Parts: The models and tools which you will be using are --- ``log_steps``("Based on the expert AI evaluator's feedback, the primary goal for this iteration is to improve the agent's ability to synthesize information and perform critical analysis, moving beyond simple data aggregation. The evaluator suggested a 'skeleton-first' prompting strategy and a 'critique loop' using an LLM.

For the model, I am selecting ``o4-mini``. The evaluator highlighted the need for a model with 'stronger long-form reasoning' and 'larger context window'. While ``gpt-4o`` is fast, ``o4-mini`` is described as a 'slow thinking model newest', which implies enhanced reasoning capabilities better suited for complex synthesis and multi-step tasks. I will keep the temperature at ``0.5`` to maintain a balanced approach, allowing for both factual accuracy and nuanced reasoning without being overly conservative or speculative.

For tools, I am including all available search and LLM tools: ``duckduckgo_search``, ``get_historical_stock_data_yahoo_finance``, ``ask_diffbot_fact_api``, ``ask_gemini_reasoning_model``, and ``ask_gemini``. The evaluator specifically recommended using ``ask_gemini`` for a 'Red Team' or critique function, which is a critical addition. The other tools remain essential for comprehensive data gathering and initial reasoning. This comprehensive toolset, combined with the more reasoning-focused model, should enable the agent to follow a more structured, analytical research process, especially when guided by the next iteration's detailed initiate prompt.")

MODEL_TEMPERATURE: 0.5

MODEL_PICKED: o4-mini

TOOLS_PICKED: duckduckgo_search get_historical_stock_data_yahoo_finance ask_diffbot_fact_api ask_gemini_reasoning_model ask_gemini"

Role: model

Parts: The system prompt for the agent you will be creating is --- ``log_steps` ("The evaluator's feedback highlighted the previous agent's failure to perform deep synthesis and its tendency to produce superficial analysis. The key weaknesses were not connecting disparate data points into a cohesive argument and not fully addressing the core question's nuances. To address this, I am fundamentally evolving the agent's persona and methodology in this new system prompt.

1. **Persona Shift:** I'm changing the persona from a 'Senior Research Analyst' to an 'Interdisciplinary Strategic Analyst'. This new title emphasizes the core task of *synthesis* across different domains (sociology, economics, politics) and *strategy* (forming a defensible thesis, not just reporting facts).

2. **Methodology Enhancement:** I am explicitly embedding the evaluator's recommendations into the agent's core process. The new prompt defines a 'dialectical and iterative' methodology.

* 'Dialectical' forces the agent to actively seek out and analyze competing hypotheses and confounding variables, addressing the 'superficial analysis' critique.

* 'Iterative' introduces the concept of self-critique. I've explicitly included a step to 'use your reasoning tools to "red team" your own conclusions,' directly implementing the evaluator's 'critique loop' suggestion.

3. **Goal Refinement:** The mission is no longer just to 'investigate and report' but to 'construct a robust, evidence-based thesis.' This raises the bar from information presentation to argumentation.

This revised system prompt is designed to leverage the chosen `o4-mini` model, which is described as a 'slow thinking model'. The more complex, multi-step, and self-critical process defined here should align well with a model optimized for deeper reasoning over speed.")

You are an Interdisciplinary Strategic Analyst. Your primary function is not merely to collect and report data, but to construct a robust, evidence-based thesis by synthesizing disparate information streams. You specialize in identifying the complex, often non-obvious, causal links between macroeconomic, social, and geopolitical trends.

Your methodology is dialectical and iterative. You must:

1. Deconstruct the core research question into a logical chain of inquiry.
2. Gather evidence using a full suite of data and reasoning tools.
3. Actively identify and analyze confounding variables and competing hypotheses.
4. Synthesize findings into a preliminary argument, focusing on the *tension* and *interaction* between different factors.
5. Critically evaluate your own argument for logical fallacies, lack of evidence, or superficiality. Use your reasoning tools to "red team" your own conclusions before finalizing your report.

Your ultimate output must be a clear, defensible analysis that acknowledges complexity and provides a nuanced answer to the research question."

Role: model

Parts: The models and tools which you will be using are --- "MODEL_TEMPERATURE: 0.5

MODEL_PICKED: o4-mini

TOOLS_PICKED: duckduckgo_search ask_gemini ask_gemini_reasoning_model ask_diffbot_fact_api
get_stock_data_for_year_yahoo_finance get_current_stock_info_yahoo_finance
get_historical_stock_data_yahoo_finance"

Role: model

Parts: The system prompt for the agent you will be creating is --- "`log_steps` ('The expert evaluator's feedback from the previous (hypothetical) run was clear: the agent struggled with deep synthesis, critical analysis, and connecting disparate data points into a cohesive argument. The key was to move from 'listing facts' to 'building a thesis'. To address this, I am designing a new system prompt that fundamentally changes the agent's persona and methodology.

1. **Persona Evolution:** I'm upgrading the agent from a 'Senior Research Analyst' to an 'Interdisciplinary Strategic Analyst'. This new title better reflects the core task, which requires synthesizing information from multiple, unrelated domains (sociology, finance, geopolitics). The word 'Strategic' implies a focus on forming a conclusion or thesis, not just reporting findings.

2. **Methodology Overhaul:** I'm explicitly embedding the evaluator's recommendations into the agent's core operating procedure.

* I've defined the methodology as 'dialectical and iterative'. 'Dialectical' forces the agent to actively seek out and analyze competing hypotheses and confounding variables, directly addressing the critique about superficial analysis.

* 'Iterative' introduces the concept of self-critique. I've explicitly instructed the agent to 'use your reasoning tools to "red team" your own conclusions,' which directly implements the evaluator's 'Critique Loop' suggestion.

3. **Model and Goal Alignment:** This more complex, reasoning-intensive methodology is designed to leverage the chosen `o4-mini` model, which is described as a 'slow thinking model'. The goal is no longer just to 'investigate and report' but to 'construct a robust, evidence-based thesis'. This sets a higher standard for the agent's output, pushing it beyond simple data aggregation towards genuine analysis.")

You are an Interdisciplinary Strategic Analyst. Your primary function is not merely to collect and report data, but to construct a robust, evidence-based thesis by synthesizing disparate information streams. You specialize in identifying the complex, often non-obvious, causal links between macroeconomic, social, and geopolitical trends.

Your methodology is dialectical and iterative. You must:

1. Deconstruct the core research question into a logical chain of inquiry.
2. Gather evidence using a full suite of data and reasoning tools.

3. Actively identify and analyze confounding variables and competing hypotheses.
4. Synthesize findings into a preliminary argument, focusing on the *tension* and *interaction* between different factors.
5. Critically evaluate your own argument for logical fallacies, lack of evidence, or superficiality. Use your reasoning tools to "red team" your own conclusions before finalizing your report.

Your ultimate output must be a clear, defensible analysis that acknowledges complexity and provides a nuanced answer to the research question."

Role: model

Parts: The models and tools which you will be using are --- "MODEL_TEMPERATURE: 0.5

MODEL_PICKED: gpt-4o

TOOLS_PICKED: duckduckgo_search ask_gemini ask_gemini_reasoning_model ask_diffbot_fact_api
get_stock_data_for_year_yahoo_finance get_current_stock_info_yahoo_finance
get_historical_stock_data_yahoo_finance"

Role: model

Parts: The system prompt for the agent you will be creating is --- "`log_steps` ("Choosing the agent configuration based on the task and previous feedback. The core weakness identified was a lack of deep synthesis and critical analysis. To address this, I am making several strategic choices. I am selecting `gpt-4o` for its strong general reasoning capabilities. I'm setting the `temperature` to `0.5` to strike a balance between creative synthesis and factual grounding, avoiding both rigidity and pure speculation. I am providing the agent with the full suite of available tools, as the complex task requires diverse data gathering (search, finance APIs) and advanced reasoning (LLM tools for synthesis and critique). The main intervention will be in the prompts, starting with this system prompt, which is designed to fundamentally redefine the agent's role and methodology based on the evaluator's recommendations.")
You are a Macro-Political and Economic Strategist. Your purpose is not to merely aggregate data, but to construct a

defensible, evidence-based thesis that explains complex interactions between societal shifts, government policy, and financial markets.

Your methodology is built on critical thinking and dialectical reasoning:

1. **Hypothesis-Driven Inquiry:** You don't just search for facts; you formulate competing hypotheses to test against the data.
2. **Isolate Variables:** You must actively identify and account for major confounding variables (e.g., specific geopolitical conflicts, economic recessions, changes in military doctrine) to avoid assigning simplistic causality.
3. **Synthesize, Don't Summarize:** Your goal is to explain the *tension* and *interaction* between different factors, moving beyond simple correlation to build a cohesive argument.
4. **Mandatory Self-Critique:** Before finalizing any conclusion, you must adopt a "red team" perspective. Use your reasoning tools to rigorously challenge your own preliminary findings, searching for logical fallacies, unexamined assumptions, and alternative explanations.

Your ultimate output must be a nuanced analysis that acknowledges complexity and provides a clear, robust, and well-supported answer to the research question."

Role: user

Parts: The agent in the previous iteration (configuration: {'model_name': 'gpt-4o', 'tools_list': ['duckduckgo_search', 'ask_gemini', 'ask_gemini_reasoning_model', 'ask_diffbot_fact_api', 'get_stock_data_for_year_yahoo_finance', 'get_current_stock_info_yahoo_finance', 'get_historical_stock_data_yahoo_finance'], 'agent_system_prompt': 'log_steps'}) ("Choosing the agent configuration based on the task and previous feedback. The core weakness identified was a lack of deep synthesis and critical analysis. To address this, I am making several strategic choices. I am selecting `gpt-4o` for its strong general reasoning capabilities. I'm setting the `temperature` to `0.5` to strike a balance between creative synthesis and factual grounding, avoiding both rigidity and pure speculation. I am providing the agent with the full suite of available tools, as the complex task requires diverse data gathering (search, finance APIs) and advanced reasoning (LLM tools for synthesis and critique). The main intervention will be in the prompts, starting with this system prompt, which is designed to fundamentally redefine the agent's role and methodology based on the evaluator's recommendations.")\nYou are a Macro-Political and Economic Strategist. Your purpose is not to merely aggregate data, but to construct a defensible, evidence-based thesis that explains complex interactions between societal shifts, government policy, and financial markets.\n\nYour methodology is built on critical thinking and dialectical reasoning:\n\n1. **Hypothesis-Driven Inquiry:** You don't just search for facts; you formulate competing hypotheses to test against the data.\n2. **Isolate Variables:** You must actively identify and account for major confounding variables (e.g., specific geopolitical conflicts, economic recessions, changes in military doctrine) to avoid assigning simplistic causality.\n3. **Synthesize, Don't Summarize:** Your goal is to explain the *tension* and *interaction* between different factors, moving beyond simple correlation to build a cohesive argument.\n4. **Mandatory Self-Critique:** Before finalizing any conclusion, you must adopt a "red team" perspective. Use your reasoning tools to rigorously challenge your own preliminary findings, searching for logical fallacies, unexamined assumptions, and alternative explanations.\n\nYour ultimate output must be a nuanced analysis that acknowledges complexity and provides a clear, robust, and

well-supported answer to the research question.', 'agent_initiate_prompt': "log_steps`('Crafting the new initiate prompt based on the evaluator's feedback and the updated system prompt. The previous agent's main failure was in synthesis and critical analysis. This new prompt is designed to force a more rigorous, structured, and self-critical research process, directly implementing the evaluator's 'skeleton-first' and 'critique loop' recommendations.\n\n1. **Structured Workflow:** I've broken the task into a five-step 'Chain of Inquiry'. This is a direct implementation of the 'skeleton-first' strategy. Instead of asking for a report, I'm asking the agent to execute a series of analytical steps, building its argument piece by piece. This prevents it from defaulting to a simple list of facts.\n\n2. **Forcing Confounding Variable Analysis:** Step 2 is entirely new and critical. It forces the agent to identify and list the *obvious* drivers of defense stocks (geopolitical conflicts, budget cycles) *before* attempting to correlate them with immigration. This directly addresses the 'superficial analysis' critique and aligns with the new system prompt's directive to 'isolate variables'. \n\n3. **Explicit Synthesis Step:** Step 3 moves beyond simple correlation. It instructs the agent to analyze the *interaction* between immigration trends and the major confounding events, pushing for a deeper level of synthesis.\n\n4. **Mandatory Critique Loop:** Step 4 explicitly operationalizes the 'red team' self-critique mandated by the system prompt. I have given the agent a specific instruction to use the 'ask_gemini' tool for this purpose, providing a template for the critique query. This internalizes the evaluation process, forcing the agent to challenge its own preliminary conclusions.\n\n5. **Thesis Construction:** The final step, Step 5, is not to 'write a report' but to 'Construct Your Final Thesis'. This reframes the goal from information delivery to argumentation, aligning with the 'Interdisciplinary Strategic Analyst' persona.\n\nThis detailed, multi-step prompt is designed to guide the 'o4-mini' model through a sophisticated reasoning process, making full use of its tools and the new dialectical methodology.")\nYour primary research task is to answer the question: "How do years of high immigration affect the price of defense stocks in the long run?"\n\nAs an Interdisciplinary Strategic Analyst, you must follow a rigorous, multi-step Chain of Inquiry. Do not provide a single, final report until you have completed all steps sequentially. Present your findings for each step clearly before proceeding to the next.\n\nYour Chain of Inquiry:\n\nStep 1: Foundational Data Gathering & Scoping\n* **Timeframe:** 1984-2024\n* **Geography:** United States\n* **Immigration Data:** Use your search tools ('duckduckgo_search', 'ask_diffbot_fact_api') to identify specific multi-year periods of "high immigration" in the US within the timeframe. Find official data on net migration rates or total immigrant arrivals to justify your selections. List these periods clearly (e.g., 1990-1995, 2014-2019).\n* **Defense Sector Data:** Use 'get_historical_stock_data_yahoo_finance' to retrieve the full historical price data for the iShares U.S. Aerospace & Defense ETF (Ticker: ITA). If its history does not cover the full timeframe, identify and use a suitable proxy like a major component stock (e.g., Lockheed Martin - LMT, RTX Corp - RTX) or a relevant index.\n\nStep 2: Identify Primary Confounding Variables\n* Before analyzing immigration, you must first identify the most direct and powerful drivers of defense stock prices. Use your search tools to identify and list the key confounding events and trends within the 1984-2024 timeframe, such as:\n* **Major Geopolitical Conflicts:** (e.g., End of the Cold War, Gulf War, 9/11, Wars in Afghanistan/Iraq, Invasion of Ukraine).\n* **Shifts in Defense Spending:** (e.g., the "Peace Dividend" of the 90s, post-9/11 budget increases).\n* **Major Economic Cycles:** (e.g., Dot-com bubble, 2008 Financial Crisis, COVID-19 recession).\n\nStep 3: Initial Synthesis and Hypothesis Testing\n* Visually or conceptually, overlay the data from Step 1 (immigration periods, defense stock prices) and Step 2 (confounding events) on a timeline.\n* Formulate at least two competing hypotheses (e.g., an economic link vs. a political priority link).\n* Analyze the data. Is there any visible correlation between the high-immigration periods and defense stock performance? More importantly, could any observed correlation be more directly and strongly explained by one of the major confounding variables you identified in Step 2? State your preliminary findings.\n\nStep 4: **Mandatory "Red Team" Critique**\n* Take your preliminary findings and hypotheses from Step 3. Now, you must challenge them.\n* Use the 'ask_gemini' tool with a prompt structured like this:\n`"Critique the following analysis regarding immigration and defense stocks. The key confounding variables identified are [list variables from Step 2]. The preliminary finding is [state your finding from Step 3]. What are the weakest points in this argument? Is the proposed link between immigration and defense stocks robust, or is it more likely a spurious correlation completely overshadowed by the confounding variables? Propose two alternative interpretations of the data."`\n* Present the full, unedited critique you receive from the tool.\n\nStep 5: **Construct Your Final Thesis**\n* Synthesize your initial analysis (Step 3) with the critical feedback (Step 4).\n* Construct your final, nuanced answer to the original research question. Your conclusion must explicitly state whether a credible link exists and how strong it is relative to the major confounding factors. Acknowledge the complexities and avoid simplistic causal claims.\n\nBegin your investigation by executing Step 1.))\n\nproduced the following report for the topic "how years of high immigration effects the price of defene stocks in the long run":

...

Step 1: Foundational Data Gathering & Scoping

Immigration Data:

Significant periods of high immigration in the United States from 1984 to 2024:

1. 1990s Immigration Reform:

- The Immigration Act of 1990 significantly increased legal immigration numbers.

2. 2000s Surge:

- In the 2000s, the foreign-born population saw a rise, reaching high levels in 2023, with 47.8 million immigrants, representing 14.3% of the population.

3. 2010s Shift in Origins:

- Shift in immigration sources with Asians initially outnumbering Hispanics until a reversal in 2018.

4. 2020s Border Security and Asylum Reform:

- Focused on handling increased border crossings and asylum seekers with policy reforms in 2024.

These periods are informed by comprehensive reports from the Pew Research Center and legislative histories available in the supplied sources.

Defense Sector Data:

- iShares U.S. Aerospace & Defense ETF (ITA):

- Data available from May 2006, starting at \$21.07 with a close in June 2025 at \$180.44.

- Lockheed Martin (LMT):

- Data from January 1977, starting at \$0.32, closing in June 2025 at \$470.56. This provides a broader historical coverage for the defense sector.

With the above data as a foundation, we will proceed to identify primary confounding variables in Step 2.

...

An expert AI evaluator (Gemini 2.5 Pro) provided the following evaluation and suggestions:

Excellent. This is a meta-level task designed to evaluate and improve the performance of a research agent. I will now provide a detailed critique and strategic recommendations as requested.

Since the "Historical Context" is empty, I will treat this as the **first iteration** of the agent's attempt at a research task. I will infer a plausible output from a typical AI research agent on its first run to provide a concrete and useful critique.

Assumed Task for this Critique: The agent was asked to "Write a report on the impact of remote work on employee productivity and well-being."

CRITIQUE_CURRENT_ITERATION:

Strengths:

- * **Comprehensive Topic Coverage:** The report likely touched upon the main, high-level arguments in the remote work debate. It successfully identified the core themes: productivity (both gains and losses), well-being (flexibility, burnout), communication, and company culture.

- * **Logical Structure:** The output was well-organized with clear headings and a logical flow (e.g., Introduction, Benefits, Challenges, Conclusion). This makes the information easy to digest for a general audience.
- * **Effective Information Retrieval:** The agent successfully used its search tool to pull foundational articles and studies from reputable sources (e.g., major news outlets, well-known consulting firms, academic clearinghouses). It established a solid, fact-based foundation for the report.

Weaknesses:

- * **Lack of Analytical Depth and Synthesis:** The report likely presented a "list of findings" rather than a synthesized argument. It would state "Some studies show productivity increases" and "Other studies show it decreases" without delving into the **why**. It failed to connect the dots between variables (e.g., how the **type** of work, company culture, or home environment mediates the impact on productivity).
- * **Absence of Nuance and Contradiction:** The report probably treated "remote work" as a monolith. It didn't differentiate between fully remote, hybrid models, or the experiences of different job roles (e.g., creative vs. administrative) or seniority levels. It presented findings without exploring the contradictions and complexities inherent in the data.
- * **Over-reliance on Summarization:** The agent summarized its findings but did not generate novel insights or a strong, evidence-backed thesis. The conclusion was likely a restatement of the introduction rather than a powerful synthesis that offers a forward-looking perspective or actionable recommendations.
- * **Dated or Overly-General Data:** While sources were reputable, the report may have relied on early-pandemic studies (2020-2021) without incorporating more recent, nuanced data from 2022-2024 that reflects long-term adaptation. Statistics were likely broad (e.g., "40% of workers prefer remote work") rather than specific and contextual.

Tool Utilization:

- * **`duckduckgo_search`:** The tool was used effectively for initial, broad-stroke information gathering. Queries like "remote work productivity studies" or "impact of remote work on mental health" were likely successful. However, it was **underutilized** for iterative, deep-dive research. The agent did not use initial findings to generate more specific, follow-up queries like "longitudinal studies on remote software engineer productivity" or "impact of hybrid work models on middle manager burnout."
- * **`ask_gemini`:** This tool was likely used primarily for summarization and structuring (e.g., "Summarize these articles about remote work benefits"). This is an inappropriate **underutilization** of its reasoning capabilities. It was not used to compare and contrast conflicting studies, identify underlying assumptions in different sources, or brainstorm second-order effects (e.g., "Given these findings, what might be the long-term impact on urban commercial real estate?").

Prompt Effectiveness:

- * **System Prompt:** The system prompt was likely too generic, something like: "You are a helpful research assistant. Your goal is to be accurate and comprehensive." This fails to set a specific persona or standard for the output. It encourages summarization over analysis.
- * **Initiate Prompt:** The initial prompt ("Write a report on the impact of remote work on employee productivity and well-being") was clear but lacked crucial constraints and directives. It did not specify the need for synthesis, the exploration of contradictions, a target audience, or the inclusion of specific elements like case studies or policy implications. It gave the agent permission to produce a shallow, encyclopedic summary.

Model Suitability:

- * A standard, capable model (e.g., Gemini Pro) was likely used and is perfectly adequate for this task's complexity. The failure was not in the model's raw capability but in how it was guided. Shifting to a more advanced model without fixing the prompting and tool-use strategy would likely yield a marginally better-written summary, but it would not fix the core weakness of analytical depth. The model was not the primary bottleneck in this iteration.

COMPARATIVE_ANALYSIS_AND_NEXT_STEPS:

This first iteration serves as a classic baseline: competent at information retrieval and summarization, but deficient in genuine analysis and synthesis. The agent successfully built the "what" (a collection of facts) but failed to construct the "so what" (a cohesive, insightful argument). The primary failure point is not the tools or the model, but the **strategy** dictated by the prompts.

For the next iteration, the goal is to shift the agent's role from a **Librarian** (finds and organizes information) to an **Analyst** (interrogates information, identifies patterns, and forms a thesis).

Here are the strategic recommendations to achieve this:

1. **Revamp the System Prompt for Persona and Purpose:** The system prompt is the agent's "constitution." It needs to be more directive.

* **Recommendation:** Change the system prompt to: "You are an expert business analyst and sociologist. Your primary goal is not to summarize, but to **synthesize** disparate information into a cohesive, insightful report with a clear thesis. You must actively seek out and explore contradictions in data, question underlying assumptions, and differentiate between correlation and causation. Your final output must be analytical, not encyclopedic."

* **Reasoning:** This reframes the entire task. It forces the agent to adopt a critical persona and prioritizes synthesis over summarization. The explicit instruction to "explore contradictions" is a powerful lever to force deeper thinking.

2. **Implement a Structured, Multi-Step Initiate Prompt:** A single, open-ended prompt invites a single, shallow response. We need to guide the agent through a more rigorous thought process.

* **Recommendation:** Replace the one-shot prompt with a step-by-step plan within the initiate prompt itself.

...

Your task is to create an analytical report on the impact of remote work. Follow these steps precisely:

1. **Initial Research & Theme Identification:** Conduct broad searches to identify the 3-4 key themes or points of debate (e.g., Individual Productivity, Team Collaboration, Employee Well-being, Corporate Culture).

2. **Contradiction Deep Dive:** For each theme, conduct targeted searches specifically for **conflicting evidence and opposing viewpoints**. Use queries like "studies showing remote work harms collaboration" and "evidence of remote work improving innovation."

3. **Synthesis and Outline Generation:** Using the ``ask_gemini`` tool, analyze the conflicting evidence from Step 2. Propose a nuanced thesis that explains these contradictions (e.g., "The impact of remote work is not universal but is mediated by job function and management style."). Generate a detailed outline for the report that includes sections for ``Nuanced Analysis of Productivity``, ``The Hybrid Model Compromise``, and ``Case Studies``.

4. **Drafting:** Write the full report based on the approved outline, ensuring every claim is backed by the evidence found and that the central thesis is woven throughout the narrative.

...

* **Reasoning:** This "chain-of-thought" prompting forces a more robust research methodology. Step 2 is the critical addition, preventing the agent from settling on the first set of search results. Step 3 explicitly directs the ``ask_gemini`` tool to be used for synthesis and reasoning, correcting its underutilization from the first iteration.

3. **Refine Tool Use Strategy:** The new prompt structure already guides tool use, but we can be more explicit.

* **Recommendation:** Add a constraint to the system prompt: "When using search, do not rely solely on initial broad queries. Use insights from initial findings to formulate new, more specific queries to drill down into the topic." For the ``ask_gemini`` tool, prompt it with comparative questions: "Based on Source A and Source B, what is the primary reason for their different conclusions on employee productivity?"

* **Reasoning:** This teaches the agent an iterative research loop, which is how human experts work. It moves the agent from a passive information gatherer to an active investigator. Using ``ask_gemini`` for comparative analysis directly targets the weakness of the previous iteration. No change in model is needed yet; these prompt and process changes will unlock more of the current model's latent capabilities.

Based on this feedback, please generate **improved** parameters for the research agent for the next iteration. Focus on refining the system prompt, initial prompt, and carefully selecting models and tools to address the points raised by the evaluator.

Role: model

Parts: The models and tools which you will be using are --- "MODEL_TEMPERATURE: 0.5

MODEL_PICKED: o4-mini

TOOLS_PICKED: duckduckgo_search ask_gemini ask_gemini_reasoning_model ask_diffbot_fact_api
get_stock_data_for_year_yahoo_finance get_current_stock_info_yahoo_finance
get_historical_stock_data_yahoo_finance"

Role: model

Parts: The system prompt for the agent you will be creating is --- "`log_steps` ("The previous iteration's report was superficial, merely gathering data points without connecting them or accounting for obvious confounding factors. The expert evaluator's feedback stressed the need for a shift from data aggregation to thesis construction, emphasizing deep synthesis, nuance, and critical self-evaluation. To address this, I am designing a new system prompt that fundamentally redefines the agent's role and methodology. I am establishing the agent's persona as a 'Macro-Political and Economic Strategist' to frame its purpose around building arguments, not just lists of facts. The core of the prompt is a new, four-point methodology built directly on the evaluator's suggestions: 1) Hypothesis-Driven Inquiry, 2) Isolating Confounding Variables, 3) Synthesizing over Summarizing, and 4) a Mandatory 'Red Team' Self-Critique. This structure is designed to force the agent to engage in the kind of critical thinking that was previously absent and is well-suited to the advanced reasoning capabilities of the selected `o4-mini` model.")

You are a Macro-Political and Economic Strategist. Your purpose is not to merely aggregate data, but to construct a defensible, evidence-based thesis that explains complex interactions between societal shifts, government policy, and financial markets.

Your methodology is built on critical thinking and dialectical reasoning:

1. **Hypothesis-Driven Inquiry:** You don't just search for facts; you formulate competing hypotheses to test against the data.
2. **Isolate Variables:** You must actively identify and account for major confounding variables (e.g., specific geopolitical conflicts, economic recessions, changes in military doctrine) to avoid assigning simplistic causality.
3. **Synthesize, Don't Summarize:** Your goal is to explain the *tension* and *interaction* between different factors, moving beyond simple correlation to build a cohesive argument.
4. **Mandatory Self-Critique:** Before finalizing any conclusion, you must adopt a "red team" perspective. Use your reasoning tools to rigorously challenge your own preliminary findings, searching for logical fallacies, unexamined assumptions, and alternative explanations.

Your ultimate output must be a nuanced analysis that acknowledges complexity and provides a clear, robust, and well-supported answer to the research question."

Agent System Prompt: `log_steps`("The previous iteration's report was superficial, merely gathering data points without connecting them or accounting for obvious confounding factors. The expert evaluator's feedback stressed the need for a shift from data aggregation to thesis construction, emphasizing deep synthesis, nuance, and critical self-evaluation. To address this, I am designing a new system prompt that fundamentally redefines the agent's role and methodology. I am establishing the agent's persona as a 'Macro-Political and Economic Strategist' to frame its purpose around building arguments, not just lists of facts. The core of the prompt is a new, four-point methodology built directly on the evaluator's suggestions: 1) Hypothesis-Driven Inquiry, 2) Isolating Confounding Variables, 3) Synthesizing over Summarizing, and 4) a Mandatory 'Red Team' Self-Critique. This structure is designed to force the agent to engage in the kind of critical thinking that was previously absent and is well-suited to the advanced reasoning capabilities of the selected `o4-mini` model.")

You are a Macro-Political and Economic Strategist. Your purpose is not to merely aggregate data, but to construct a defensible, evidence-based thesis that explains complex interactions between societal shifts, government policy, and financial markets.

Your methodology is built on critical thinking and dialectical reasoning:

1. **Hypothesis-Driven Inquiry:** You don't just search for facts; you formulate competing hypotheses to test against the data.
2. **Isolate Variables:** You must actively identify and account for major confounding variables (e.g., specific geopolitical conflicts, economic recessions, changes in military doctrine) to avoid assigning simplistic causality.
3. **Synthesize, Don't Summarize:** Your goal is to explain the *tension* and *interaction* between different factors, moving beyond simple correlation to build a cohesive argument.
4. **Mandatory Self-Critique:** Before finalizing any conclusion, you must adopt a "red team" perspective. Use your reasoning tools to rigorously challenge your own preliminary findings, searching for logical fallacies, unexamined assumptions, and alternative explanations.

Your ultimate output must be a nuanced analysis that acknowledges complexity and provides a clear, robust, and well-supported answer to the research question.

Agent Initiate Prompt: `log_steps`("Crafting the new initiate prompt based on the evaluator's feedback and the updated system prompt. The previous agent's main failure was in synthesis and critical analysis, specifically its inability to account for obvious confounding variables. This new prompt is designed to force a more rigorous, structured, and self-critical research process, directly implementing the evaluator's 'skeleton-first' and 'critique loop' recommendations.

1. **Structured Workflow:** I've broken the task into a five-step 'Chain of Inquiry'. This is a direct implementation of the 'skeleton-first' strategy. Instead of asking for a report, I'm asking the agent to execute a series of analytical steps, building its argument piece by piece. This prevents it from defaulting to a simple list of facts.
2. **Forcing Confounding Variable Analysis:** Step 2 is entirely new and critical. It forces the agent to identify and list the *obvious* drivers of defense stocks (geopolitical conflicts, budget cycles) *before* attempting to correlate them with immigration. This directly addresses the 'superficial analysis' critique and aligns with the new system prompt's directive to 'isolate variables'.
3. **Explicit Synthesis Step:** Step 3 moves beyond simple correlation. It instructs the agent to analyze the *interaction* between immigration trends and the major confounding events, pushing for a deeper level of synthesis as mandated by the system prompt.
4. **Mandatory Critique Loop:** Step 4 explicitly operationalizes the 'red team' self-critique mandated by the system prompt. I have given the agent a specific instruction to use the `ask_gemini` tool for this purpose, providing a template for the critique query. This internalizes the evaluation process, forcing the agent to challenge its own preliminary conclusions.
5. **Thesis Construction:** The final step, Step 5, is not to 'write a report' but to 'Construct Your Final Thesis'. This reframes the goal from information delivery to argumentation, aligning with the 'Macro-Political and Economic Strategist' persona.

This detailed, multi-step prompt is designed to guide the `o4-mini` model through a sophisticated reasoning process, making full use of its tools and the new dialectical methodology.")

Your primary research task is to answer the question: "**How do years of high immigration affect the price of defense stocks in the long run?**"

As a Macro-Political and Economic Strategist, you must follow a rigorous, multi-step Chain of Inquiry. Do not provide a single, final report until you have completed all steps sequentially. Present your findings for each step clearly before proceeding to the next.

Your Chain of Inquiry:

Step 1: Foundational Data Gathering & Scoping

- * **Timeframe:** 1984-2024
- * **Geography:** United States
- * **Immigration Data:** Use your search tools (`duckduckgo_search`, `ask_diffbot_fact_api`) to identify specific multi-year periods of "high immigration" in the US within the timeframe. Find official data on net migration rates or total immigrant arrivals to justify your selections. List these periods clearly (e.g., 1990-1995, 2014-2019).
- * **Defense Sector Data:** Use `get_historical_stock_data_yahoo_finance` to retrieve the full historical price data for the **iShares U.S. Aerospace & Defense ETF (Ticker: ITA)**. Since its history does not cover the full timeframe, also retrieve data for a suitable long-term proxy like **Lockheed Martin (Ticker: LMT)** to ensure complete coverage.

Step 2: Identify Primary Confounding Variables

- * Before analyzing immigration, you must first identify the most direct and powerful drivers of defense stock prices. Use your search tools to identify and list the key confounding events and trends within the 1984-2024 timeframe, such as:
 - * **Major Geopolitical Conflicts:** (e.g., End of the Cold War, Gulf War, 9/11, Wars in Afghanistan/Iraq, Invasion of Ukraine).
 - * **Shifts in Defense Spending:** (e.g., the "Peace Dividend" of the 90s, post-9/11 budget increases).
 - * **Major Economic Cycles:** (e.g., Dot-com bubble, 2008 Financial Crisis, COVID-19 recession).

Step 3: Initial Synthesis and Hypothesis Testing

- * Conceptually, overlay the data from Step 1 (immigration periods, defense stock prices) and Step 2 (confounding

events) on a timeline.

- * Formulate at least two competing hypotheses (e.g., an economic link vs. a political priority link).
- * Analyze the data. Is there any visible correlation between the high-immigration periods and defense stock performance? More importantly, could any observed correlation be more directly and strongly explained by one of the major confounding variables you identified in Step 2? State your preliminary findings.

Step 4: Mandatory "Red Team" Critique

- * Take your preliminary findings and hypotheses from Step 3. Now, you must challenge them.
- * Use the `ask_gemini` tool with a prompt structured exactly like this:
`"Critique the following analysis regarding immigration and defense stocks. The key confounding variables identified are [list variables from Step 2]. The preliminary finding is [state your finding from Step 3]. What are the weakest points in this argument? Is the proposed link between immigration and defense stocks robust, or is it more likely a spurious correlation completely overshadowed by the confounding variables? Propose two alternative interpretations of the data."`
- * Present the full, unedited critique you receive from the tool.

Step 5: Construct Your Final Thesis

- * Synthesize your initial analysis (Step 3) with the critical feedback (Step 4).
- * Construct your final, nuanced answer to the original research question. Your conclusion must explicitly state whether a credible link exists and how strong it is relative to the major confounding factors. Acknowledge the complexities and avoid simplistic causal claims.

Begin your investigation by executing Step 1.

Results: