mining

MDPI

Article

MINDS: A Modular Multi-Agent Decision-Support Framework for Dynamic Strategic Mine Planning

Ricardo Nunes, Nathalie Risso * and Moe Momayez

School of Mining Engineering and Mineral Resources, University of Arizona, Tucson, AZ 85721, USA; ricardonunes@arizona.edu (R.N.); mmomayez@arizona.edu (M.M.)
* Correspondence: nrisso@arizona.edu

Abstract

Strategic Mine Planning (SMP) creates the long-term economic baseline for mining operations, yet economic variability necessitates Dynamic Mine Planning (DMP) to rapidly stress-test those financial assumptions. Currently, this capability is hindered by fragmented software ecosystems that require manual data handoffs, slowing iteration and breaking the audit trail between market data and valuation models. While Generative AI affords an opportunity to automate these workflows, its adoption in the mining industry is stalled by concerns over data quality and the risk of uncritical acceptance of automated outputs. Addressing these challenges, this paper describes the Mine Intelligence and Decision Support (MINDS) framework. We present MINDS as a modular reference architecture that uses Large Language Model (LLM) agents to orchestrate the economic evaluation process while maintaining strict engineering oversight. The system integrates a conversational interface with a multi-agent assessment layer that acts as an adversarial review, assessing price assumptions against market intelligence before generating economic valuation scenarios. A proof-of-concept using the Marvin copper benchmark evaluates the framework, demonstrating automated request-to-report orchestration, execution stability with an average debate latency of 10.69 s and a transparent decision audit trail. These findings show that MINDS can systematize economic scenario analysis without sacrificing the governance and verification required for definitive feasibility studies.

Keywords: strategic mine planning; dynamic mine planning; large language models; AI agents; multi-agent debate; decision audit trail; decision support systems

Check for updates

Academic Editor: Roohollah Shirani Faradonbeh

Received: 19 February 2026
Revised: 18 March 2026
Accepted: 25 March 2026
Published: 2 April 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

At its core, SMP establishes the long-term economic objectives of a mining operation over horizons typically exceeding five years, and it is fundamentally a static framework. DMP is not a competing timeframe, but a different capability altogether: the organizational capacity to revise that framework as markets shift or new data arrives. Traditional workflows are usually fragmented across disconnected software tools, requiring manual data handoffs that break the audit trail between market assumptions and the final mine design. The financial consequence of this planning gap is substantial; recent analyses reveal that mining companies missed production targets by USD 64 billion over the past five years. On average, organizations underperformed their own forecasts by 2.4% annually, with slow, manual planning cycles identified as the primary root cause [1].

These integration challenges are becoming more consequential as demand for critical minerals accelerates. Global demand for critical minerals is projected to triple by 2030

and quadruple by 2040 [2], placing immense pressure on mining companies to accelerate digital transformation amid widening supply–demand imbalances [3]. SMP sits at the center of this pressure, requiring the synthesis of geological models, market forecasts, and optimization algorithms. Although modern mines produce large amounts of data, the success of DMP depends on high-quality, standardized inputs to ensure accurate economic assessments [4,5]. However, seamlessly exchanging reliable data between these domains remains difficult, often leaving workflows dependent on manual intervention.

In recent years, a new paradigm has emerged to bridge these gaps: Generative Artificial Intelligence (AI). Unlike traditional numerical algorithms, LLMs are designed to process human language and reasoning. This capability allows them to function as agents that support human engineers by interpreting natural language instructions, planning execution sequences, and actively driving external software tools. These capabilities offer a pathway to orchestrate multi-step workflows across heterogeneous systems [6–8]. Yet deploying such agents in high-consequence environments demands governance structures that strictly maintain human decision authority [9].

In response, mining suppliers have begun deploying AI-powered systems to streamline technical information access. ABB's GMD Copilot enables natural-language interaction with documentation integrated into monitoring platforms [10], while Airth provides AI-driven decision support specifically for strategic mine scheduling [11]. However, these commercial platforms typically operate as closed, proprietary systems that limit extensibility and integration with third-party solvers. Parallel to these industrial advancements, academic research has investigated LLM-based agents for mining applications, enhancing transparency through natural language reasoning [12]. A clear distinction exists in the literature: while implementations have targeted operational contexts like autonomous vehicle control [13] and truck dispatch scheduling [14], SMP workflows lack comparable attention. Specifically, the field offers no comprehensive framework (a standardized digital backbone designed to link isolated software tools into a cohesive system) that supports the automated orchestration of decision pipelines to integrate conversational interfaces, multi-agent economic assessment, and human approval gates.

The primary obstacle manifests at the integration boundaries between these heterogeneous technologies. Realizing the potential of DMP requires overcoming three fundamental barriers:

1. **The Integration Barrier:** Fragmented workflows force manual data transfer between geology, markets, and solvers, preventing the rapid iteration required by volatile commodity cycles.
2. **The Stochastic-Deterministic Alignment:** Capital justification demands deterministic reproducibility, yet LLMs are inherently probabilistic. Aligning these opposing paradigms requires architectural constraints that enforce consistent, defensible engineering outcomes.
3. **Lack of Decision Auditability:** Autonomous agents often operate as opaque closed systems. SMP requires more than optimized geometry; it demands a transparent audit trail linking the final design back to its originating economic assumptions.

To address these limitations, this paper proposes the MINDS framework. While commercial tools such as ABB's GMD Copilot and Airth confirm growing industry appetite for AI-assisted planning, their closed and proprietary nature limits their extensibility and integration with third-party solvers, leaving open-architecture solutions for SMP largely unexplored. This modular architecture is developed to fill the gap between SMP and DMP workflows. It applies Generative AI to SMP by using LLMs as autonomous agents that interpret natural language intent to orchestrate technical workflows. The system relies on a four-layer design comprising Optimization and Data, Economic Assessment, Interface and Orchestration, and Output Generation. This structure preserves the sep-

aration of concerns while supporting end-to-end integration. Scenario initiation occurs via natural language, triggering automated parameter extraction, multi-agent economic assessment integrating real-time market intelligence, pit optimization upon explicit user approval, and automated report generation. The framework is designed for technology-agnostic component substitution through interface abstraction. Capabilities are demonstrated with a case study using the Marvin copper benchmark [15], executing complete request-to-report workflows.

The primary contributions of this paper are as follows:

- A **modular reference architecture** for DMP that enables technology-agnostic substitution, allowing engineers to swap optimization solvers, pricing models, or interface components independently without disrupting the surrounding workflow.
- A **conversational workflow orchestration system** that automates the complete scenario lifecycle from natural language engineering requests through pit optimization to report generation, addressing the workflow fragmentation caused by manual data handoffs across disconnected planning tools.
- A **multi-agent assessment module** that exploits the role-impersonation capability of LLMs to simulate adversarial economic review, enforcing a structured critique of commodity price assumptions before optimization is triggered, rather than accepting user inputs uncritically as single-model systems tend to do.

2. Background and Related Work

Recent reviews establish LLM-driven agents as an emerging paradigm for mining decision support [16,17]. Building on that foundation, this section focuses on two areas essential to MINDS: workflow orchestration challenges and multi-agent frameworks for economic analysis.

2.1. Workflow Orchestration

Systematic reviews identify workflow fragmentation as a critical SMP bottleneck [16,17]. This issue is largely organizational: manual parameter coordination across disconnected tools creates friction when timely scenario iteration is needed to respond to market volatility [18,19].

Current research has demonstrated the utility of LLMs in executing discrete operational tasks, including autonomous vehicle control [13], geotechnical site planning [20], and safety monitoring [4]. Data integration efforts have leveraged knowledge graphs (structured databases mapping relationships between assets and records) for maintenance [21], hybrid semantic search (improving retrieval by understanding intent over keywords) [5], and multi-level Industrial Internet of Things (IIoT) architectures (networks integrating sensors and equipment for data collection) [22]. Parallel advancements in multi-agent reinforcement learning (where software agents learn optimal decisions through trial and error) focus on logistics like truck dispatching [14,23–25]. However, these operational successes have not yet translated into unified SMP frameworks.

2.2. Multi-Agent Frameworks

In capital-intensive decision-making, unverified assumptions often lead to strategic failures [1,19]. A specific risk in AI-assisted planning is *sycophancy*, where models prioritize user agreement over factual accuracy, allowing for unrealistic production rates or price forecasts to pass unchallenged [26,27]. To counter this, multi-agent debate frameworks introduce an automated technical audit. By assigning conflicting roles, such as an expansion advocate versus a risk auditor, these systems force a structured disagreement that mimics professional peer review. Research confirms that this adversarial process significantly

reduces errors compared to single-model outputs, preventing the "echo chamber" effect that endangers long-term infrastructure commitments [26,27].

The effectiveness of this coordination pattern is well-established in other high-stakes domains. Financial trading systems have deployed "Bull-Bear" debate patterns to model market volatility [28], while actor–critic–expert architectures have achieved 87% accuracy in query resolution by embedding critique into the reasoning loop [29]. Similarly, zero-shot architectures demonstrate that integrating technical and macroeconomic signals can outperform conventional time-series models in risk-sensitive contexts [30].

These precedents are critical for SMP, where market volatility often forces expensive rework of multi-year schedules [31]. However, a gap remains in the current mining literature: while recent work has successfully combined price signals with automated news extraction [32], existing agents typically function as standalone tools rather than integrated validation stages [33,34]. Consequently, the industry lacks a framework that enforces multi-agent economic critique as a mandatory gateway before pit optimization—a governance pattern that aligns with the rigor required by human investment committees.

3. MINDS Framework Architecture

MINDS aims to transform SMP from fragmented manual workflows into integrated conversational pipelines. Instead of requiring manual data transfer across disconnected tools, the framework automates the entire lifecycle through natural language requests. This automated process executes three critical functions: multi-agent assessment incorporating market intelligence, pit optimization subject to explicit human approval, and the generation of audit trails. To support this integration, the framework relies on a modular architecture where an orchestration engine manages component transitions.

3.1. Architectural Overview

The MINDS architecture comprises four functional layers (Figure 1):

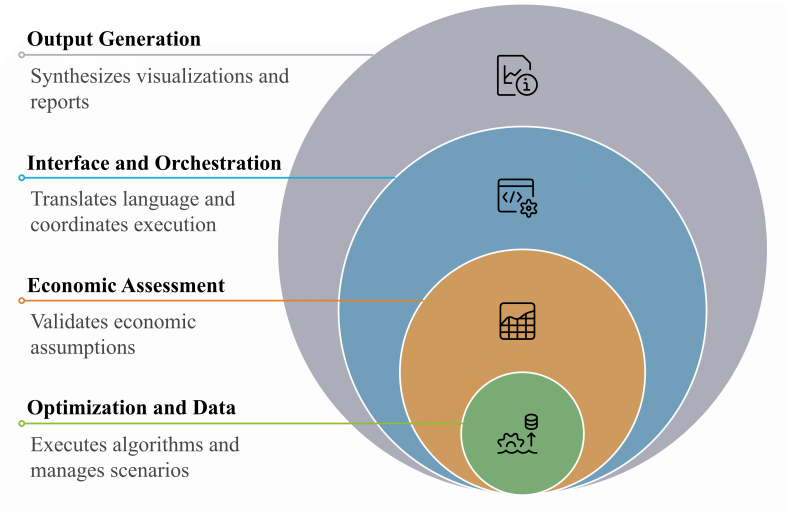


Figure 1. MINDS framework architecture stack showing four functional layers enabling workflow automation.

1. *Optimization and Data Layer:* Executes pit optimization algorithms and manages scenarios with automated traceability.
2. *Economic Assessment Layer:* Evaluates economic assumptions through multi-agent debate integrating market intelligence.
3. *Interface and Orchestration Layer:* Translates natural language into structured workflows and coordinates multi-step execution.

4. *Output Generation Layer:* Synthesizes visualizations and reports for decision support.

Figure 2 illustrates how the framework automates the data transfer between typically disconnected tasks. The central orchestrator manages the exchange of parameters between the market analysis components and the optimization engine, eliminating the need for manual file handling or spreadsheet transfers. Solid arrows indicate the automated execution flow, while double-headed arrows show how data is standardized and persisted at every step. This structure allows for the engineer to focus on the decision signals rather than the mechanics of moving data between software packages.

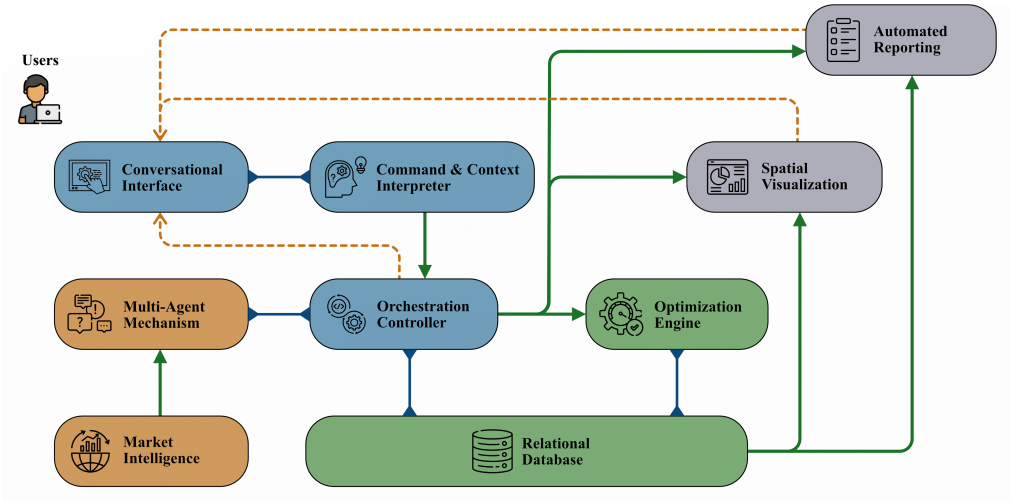


Figure 2. Component interaction diagram showing the orchestration-centric workflow and user interaction points. Solid arrows indicate the automated sequential execution flow. Double-headed arrows represent bidirectional data exchange. Dashed arrows represent user-initiated feedback, allowing engineers to review outputs.

Table 1 details the nine core components implementing these layers. By separating architectural roles from their reference implementations, the design enforces that system integration points remain functional, even if specific technologies are substituted.

Table 1. Core components, showing architectural separation between roles and implementations.

Layer	Component (Role)	Reference Implementation
Interface and Orchestration	**Conversational Interface** *multi-modal input* **Command & Context Interpreter** *Parameter extraction* **Orchestration Controller** *Execution coordination*	**Telegram Bot API** *Voice/text support* **OpenAI Whisper + GPT-4o-mini** *Speech context extraction* **Python State Controller** *Async task management*
Economic Assessment	**Multi-Agent Mechanism** *Assumption Evaluation* **Market Intelligence** *Data integration*	**LangGraph + OpenAI** *Parallel agent orchestration* **yfinance + NewsAPI** *Historical prices + news sentiment*
Optimization and Data	**Optimization Engine** *Pit limit calculation* **Relational Database** *Scenario storage*	**NetworkX/SciPy** *Max-closure algorithms* **PostgreSQL 16** *Database with automated triggers*
Output Generation	**Automated Reporting** *Analytical synthesis* **Spatial Visualization** *Outcome rendering*	**ReportLab** *Automated PDF reports* **Plotly** *Interactive 3D models*

3.2. Optimization and Data Layer

The Optimization and Data Layer integrates algorithmic execution with automated scenario management. The Optimization Engine implements pit limit calculation using the Lerchs–Grossmann (LG) algorithm [35]. This algorithm was selected as the default reference implementation because it is the established industry standard for determining the ultimate pit; unlike heuristic methods, LG guarantees a mathematically optimal solution, ensuring that any variation in the final mine plan is driven strictly by the AI's economic strategy rather than solver noise. However, the framework treats this solver as a modular service: by enforcing standardized input/output contracts, the system allows for engineers to swap in alternative solvers without modifying the surrounding orchestration logic.

The framework employs a relational database that unifies heterogeneous data types. This system stores the block model as structured spatial data while simultaneously preserving semi-structured debate logs as flexible JSON objects (i.e., a human-readable set of named fields). Figure 3 illustrates the high-level data architecture, separated into three functional groups: *core inputs* defining scenario assumptions, *decision orchestration* components that preserve both the debate rationale and execution logs, and *strategic outcomes* capturing the final mining geometry. This architecture enforces that both the *Optimization Engine Service* and *Multi-Agent Debate* logic operate within a traceable context. Consequently, users can query the database to reconstruct the complete decision-making process, ensuring full end-to-end traceability from initial intent to final design.

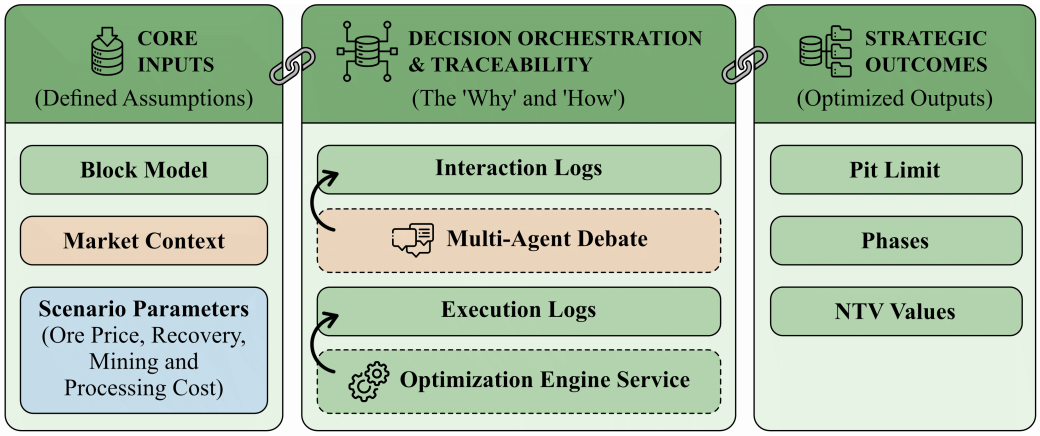


Figure 3. Conceptual data architecture implementing the audit trail. The central layer orchestrates two distinct services: Multi-Agent Debate for economic evaluation, and the Optimization Engine for physical planning. Both services write to persistent logs, ensuring that all strategic outcomes are traceable back to their originating core inputs.

3.3. Economic Assessment Layer

The Economic Assessment Layer functions as a mandatory gateway, ensuring that economic assumptions undergo rigorous stress-testing before the system commits to computationally expensive optimization. This layer transforms raw user price proposals into evidence-based deliberation by integrating real-time market intelligence with parametric cost models.

We selected a multi-agent design to mitigate confirmation bias. Unlike single-model chatbots that tend to validate user inputs regardless of market reality, this architecture forces an objective review. By assigning conflicting roles to independent agents, specifically mandating one to identify opportunities and another to audit risks, the framework enforces an adversarial review that mirrors the deliberative rigor of a human investment committee. Although this reference implementation relies on LLM-based agents to demonstrate autonomous reasoning, the architecture defines this assessment function as a modular

interface. Consequently, the validation logic remains interchangeable and can be substituted with deterministic econometric algorithms or human-expert review panels without requiring modifications to the broader orchestration workflow.

To establish a factual baseline for this debate, the framework deploys a Market Intelligence Service that aggregates data from two distinct streams: a quantitative stream retrieves three-year historical copper price data via the Yahoo Finance API to calculate volatility and momentum statistics, while a qualitative stream sources industry news from the preceding 30 days through NewsAPI. Commodity-specific keyword queries are applied at retrieval, followed by post-filtering to retain only copper-related articles, producing a context that includes price percentile positioning, volatility statistics, 30- and 90-day trends, and a sentiment score derived from bullish and bearish signal counts. As a modular component, the Market Intelligence Service can be substituted with alternative data providers or internal pricing models to meet the data governance requirements of a specific deployment.

The evaluation process is executed by three specialized agents (Figure 4): The *Upside Agent* advocates for opportunity, utilizing sentiment data to determine if conservative proposals ignore potential market rallies. In contrast, the *Downside Agent* acts as a risk auditor, citing historical volatility to challenge overly optimistic projections. Finally, the *Strategy Agent* synthesizes these opposing perspectives into a balanced recommendation.

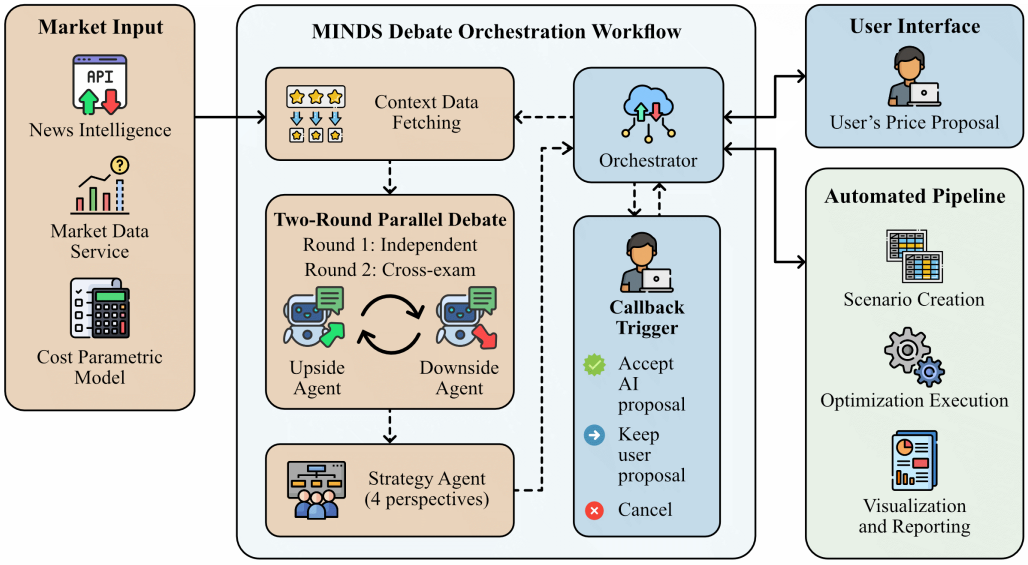


Figure 4. Multi-agent debate orchestration workflow. The process is initiated by the user's price proposal (**right top**) and follows a sequential path: market inputs are retrieved (**left top**), a bounded two-round debate between the Upside and Downside agents is executed, and the Strategy Agent synthesizes a recommendation. The circular arrows within the debate box represent the two structured rounds of exchange. The workflow concludes at the Callback Trigger, where explicit human approval routes the decision to the Automated Pipeline (**right bottom**) for scenario creation, optimization, and reporting.

Interaction between these agents follows a structured two-round protocol designed to mitigate confirmation bias. In the first round, agents perform independent analyses to prevent premature consensus. The second round introduces cross-examination, where agents critique each other's findings to surface logical inconsistencies.

To enforce financial realism, the system applies a parametric cost escalation model. In the mining industry, higher commodity prices often correlate with inflated input costs, such as fuel, labor, and consumables. The system quantifies this relationship using a cost–price elasticity formula. For example, with an elasticity of $\varepsilon = 0.73$, a 10% increase in copper price implies a 7.3% increase in Operating Expenses (OpEx), as captured by (1).

$$\mathrm{OpEx}(P) = \mathrm{OpEx}_0 \left(\frac{P}{P_0} \right)^{\varepsilon} \quad (1)$$

where P_0 is the baseline price and OpEx_0 is the baseline cost. The elasticity value of $\varepsilon = 0.73$ is an illustrative parameter reflecting the tendency for fuel, labor, and consumable costs to rise alongside commodity prices in mining operations. In this proof of concept, the model is applied to OpEx adjustments to demonstrate how economic assumptions can be stress-tested prior to optimization. This mechanism prevents the assumption that revenue increases flow directly as pure margin.

3.4. Interface and Orchestration Layer

The Interface and Orchestration Layer enables natural language interaction with the planning system through conversational queries and commands. Users access market data, query stored scenarios, and initiate workflow execution through text or voice inputs (Figure 5). Three components operate in sequence: the User Interface for multi-modal capture, Intent Understanding for semantic parsing and command routing, and Workflow Orchestration for execution coordination across the scenario evaluation, optimization, and reporting stages.

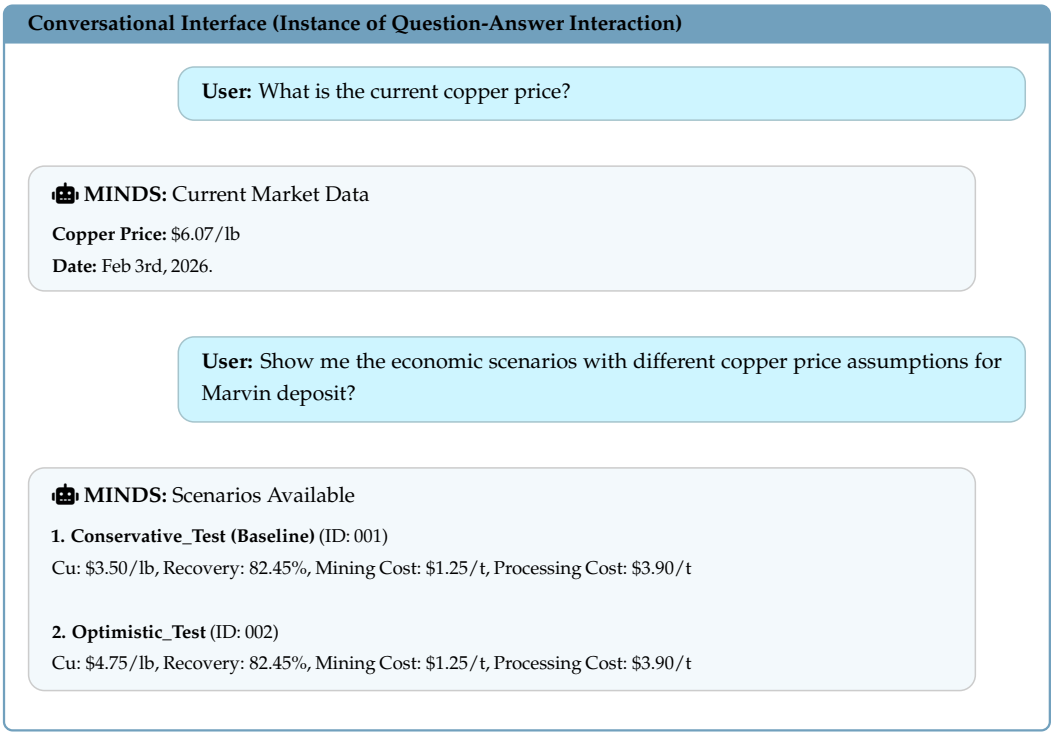


Figure 5. Natural language interaction demonstrating conversational query handling (market data retrieval, scenario listing) through text-based dialogue.

These components form a unified pipeline that normalizes multi-modal inputs. Voice transcriptions and typed text feed into a shared intent classifier, reducing interface variability. To enforce accuracy, the system uses confidence-based gating. It triggers automated execution only when high confidence is detected, prompting the user for clarification otherwise. Simultaneously, parameter extraction algorithms convert unstructured text into valid data schema.

Decision governance relies on a sequential logic structure that strictly enforces human oversight. For example, once the assessment is complete, the system enters an approval-pending mode, where it interprets inputs exclusively as decision signals. This architecture

enforces that the system commits computational resources and logs actions only after explicit authorization. Throughout this process, intermediate status updates provide transparency, mitigating the opacity often found in asynchronous multi-step workflows.

3.5. Output Generation Layer

The Output Generation Layer converts optimization results into decision-ready outcomes through interactive 3D visualizations and structured analytical reports. Visual outputs present three-dimensional block models, optimized pit shells, and phase sequences, allowing for engineers to inspect pit geometry, phase boundaries, and extraction progression in a spatially intuitive manner. Analytical reports are supported by concise interpretive commentary that highlights technical and economic implications. All outputs are derived from a common set of planning assumptions and optimization results, ensuring consistency, traceability, and alignment across visualization, analysis, and reporting throughout the planning workflow.

4. Case Study: Marvin Benchmark Demonstration

This section demonstrates MINDS capabilities through a case study using the Marvin copper benchmark, a widely-adopted synthetic block model from the Minelib repository [15]. The Marvin dataset is a synthetic copper–gold block model originally designed to reproduce the structural characteristics of a hydrothermal sulfide deposit, comprising 53,271 blocks with grade, tonnage, and geospatial precedence attributes. Its scale and deposit structure are representative of open-pit planning problems encountered in practice, and it has been widely adopted as a standard testbed for evaluating optimization algorithms and production scheduling methods [36,37].

Block economics are defined by the Net-Ton-Value (NTV), representing the revenue per tonne minus operational costs. The NTV for each block is calculated as

$$NTV = (P_{Cu} \times 2000 \times \frac{g}{100} \times R) - (C_{mine} + C_{proc}) \quad (2)$$

where P_{Cu} is the copper price (\$/lb), g is the copper grade (%), R is the total recovery, and C_{mine} and C_{proc} are the mining and processing costs per tonne, respectively.

Table 2 summarizes the dataset and the fixed parameters used for this demonstration. Mining and processing costs are treated as simplified unit costs applied uniformly across all blocks, consistent with the proof-of-concept scope of this study. The modular architecture supports more detailed cost formulations, including the separation of fixed and variable components, in future implementations.

Table 2. Marvin benchmark dataset characteristics and case study configuration.

Attribute	Value
Block Model Characteristics	
Total Blocks	53,271
Block Dimensions	30 × 30 × 30 m
Commodities	Copper (Cu), Gold (Au)
Precedence Constraints	45° slope angle, 8 levels
Dataset Type	Synthetic benchmark
Model Parameters (for NTV * Calculation)	
Primary Commodity	Copper (Cu only)
Total Recovery	82.45%
Mining Cost	\$1.25/t
Processing Cost	\$3.90/t

* NTV: The deterministic economic value of a block, calculated using (2).

4.1. Case Study Design

This case study, conducted on 3 February 2026 (copper spot price: $6.07/lb, retrieved via the Yahoo Finance API), employs two complementary approaches to demonstrate MINDS' capabilities across different dimensions of system performance. The demonstration focuses on an open-pit, single-commodity deposit, consistent with the Marvin benchmark configuration.

The evaluation adopts two performance criteria aligned with the proof-of-concept scope of this work. The first is *orchestration stability:* whether the multi-agent debate mechanism completes reliably and produces coherent economic recommendations across repeated invocations. The second is *decision coherence*: whether the agents contextualize user price proposals against market intelligence rather than accepting them uncritically. These criteria are intentionally scoped to integration and workflow behavior. The effectiveness of multi-agent debate in reducing reasoning errors relative to single-model outputs has been established in prior work [26,27], and this study adopts that validated architectural pattern rather than reproducing its benchmarking.

First, the end-to-end workflow demonstration shows the complete pipeline's capability to translate a natural language request into executable optimization deliverables. Through the conversational interface, a scenario was initiated using real-time market data. The system automatically applied standardized operational parameters: 82.45% total recovery, $1.25/t mining cost, and $3.90/t processing cost. These parameters were used to calculate Net-Ton-Value and determine block economic viability.

Second, to assess the robustness of the multi-agent orchestration layer, we executed 20 repeated debate runs across three distinct price levels: conservative ($5.00/lb, 7 trials), moderate ($5.80/lb, 7 trials), and optimistic ($6.50/lb, 6 trials). Each run was executed sequentially under stable market conditions without manual intervention. All agent prompts were fixed across runs, and inference was configured with a temperature of 0.3, intentionally preserving a degree of stochasticity to reflect realistic deployment conditions. Orchestration latency and the qualitative consistency of AI-generated price assessments were recorded for each trial.

4.2. End-to-End Workflow Demonstration

This demonstration executes a complete "request-to-execute" workflow, showing the framework's ability to autonomously link SMP stages. Initiated via a single natural language command, the system orchestrates a continuous sequence of tasks, scrutinizing economic assumptions through multi-agent debate, and, upon receiving human approval, executing the LG optimization sequence. This process verifies that the architecture can successfully bridge the gap between unstructured intent and structured engineering outputs without requiring manual data transfer.

As shown in Figure 6, the pipeline was triggered by a user request to evaluate the current market price of $6.07/lb. The system automatically retrieved this context, triggered the multi-agent debate, and synthesized a strategy that recommended a −2.8% adjustment to $5.90/lb, buffering against peak-market volatility while acknowledging strong bullish sentiment. Upon receiving human approval, the system automatically executed the optimization workflow. The assessment confirmed end-to-end integration: the system classified 11,405 blocks into two mining phases and generated interactive visualizations without manual file transfers. Figure 7 shows the automatically generated phase geometries produced by the optimization workflow, providing an immediate 3D verification outcome for engineering review.

Multi-Agent Assessment (Abridged Transcript)

User: Create a new scenario with current copper price using default parameters

MINDS: Market Context Retrieved

Spot: $6.07/lb (96th percentile).
Sentiment: Bullish (+20 score, 29 bullish / 9 bearish signals).

MINDS: Agent Debate Execution

Upside Agent: "Current price at 96th percentile near 3-year high. [...] Copper rally and supply constraints support $6.25/lb."
Downside Agent: "Revenue rises 51.8%, yet OpEx inflates 35.6% (ε=0.73), yielding only 16.2% true margin gain. [...] Recommend $5.74/lb."
Cross-Exam: Upside argues bullish fundamentals underappreciated; Downside warns 9.0% volatility at peak creates risk.

MINDS: Strategy Recommendation

Adjust to $5.90/lb (−2.8%).
Rationale: Price at 96th percentile warrants caution. [...] OpEx inflation and volatility justify buffer below spot.

User: (Human Decision)
✓ Accept AI Recommendation

MINDS: Starting full pipeline workflow execution at $5.90/lb

Lerchs–Grossmann optimization complete.
Deliverables: 3D Block Model, Pit Shell Surface, Phases Sequence...

Figure 6. Abridged transcript of multi-agent assessment showing conversational workflow from the 3 February 2026 demonstration.

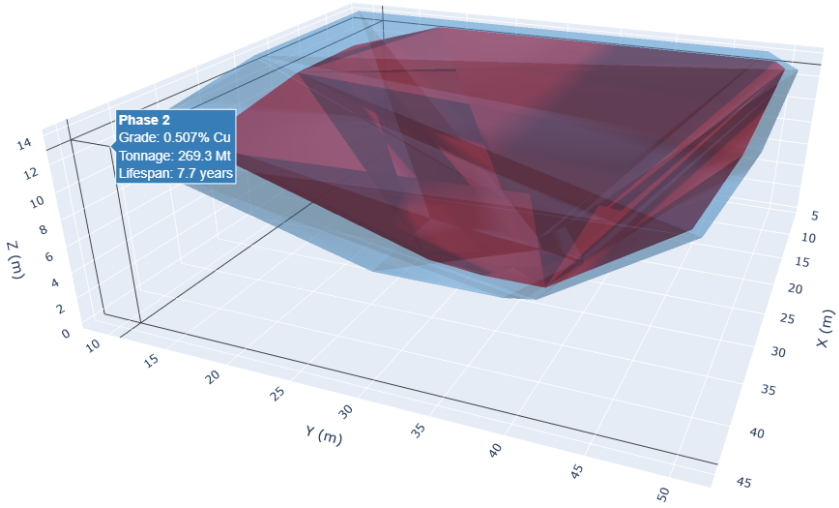


Figure 7. Illustrative MINDS output produced after optimization and phase assignment, visualizing pit phases and their summary statistics.

Audit Trail: Database audit trails preserve the full decision context, logging user prompts, market data inputs, valuation parameters, and intermediate agent decisions as structured records. Table 3 illustrates this traceability, reconstructing the execution flow from initial user intent to final approved parameters. By using a normalized relational schema, the system enforces that audit queries can track decision history across accumulated scenarios, revealing exactly when and why the AI recommended adjusting the user's proposal.

Table 3. Decision traceability demonstrating a complete audit trail. The log captures the user's at-market proposal ($6.07) matching the 96th percentile spot price, documenting the rationale for the AI's recommended adjustment ($5.90) to buffer against volatility.

Attribute	Value	Time
User Intent and Market Context		
User Proposal	$6.07/lb (at spot)	13:37:11
Spot Price	$6.07/lb	13:37:15
Hist. Percentile	96th	13:37:15
Sentiment	Bullish (+20, 29 bullish/9 bearish)	13:37:18
Multi-Agent Debate Rationale Logs		
Upside Agent	"Near historical high with copper rally. Supply constraints support $6.25/lb [...]"	13:37:22
Downside Agent	"At 96th percentile, OpEx inflation (ε = 0.73) yields only 16.2% true margin gain. Volatility 9.0% warrants buffer [...] recommend $5.74/lb"	13:37:24
Strategy Agent	"Recommend $5.90/lb (−2.8%): balances bullish sentiment with volatility buffer below peak [...]"	13:38:28
Human Decision and Optimization Outcomes		
AI Recommendation	$5.90/lb (−2.8%)	13:38:28
User Decision	Accept AI Recommendation	13:38:35
Final Price	$5.90/lb	13:38:45
Pit Tonnage	693,182,180 t	13:39:47
Pit Value	$14,496,677,999	13:39:47
Blocks In-Pit	11,405/53,271	13:39:47

4.3. Stability and Coherence Trials

Quantitative performance metrics from 20 controlled trials present the framework across two dimensions: computational predictability (Table 4) and economic decision coherence (Table 5). These metrics confirm that the system maintains consistent production characteristics while demonstrating rational, context-aware decision logic.

Computational Stability: The results indicate consistent temporal stability despite stochastic LLM generation. Debate latency averaged 10.69 s with a coefficient of variation of 14.2%, supporting expectations for sub-12-second execution. As shown in Figure 8, parallel agent execution accounts for 47% of workflow time (Round 1 + Round 2 $\approx$ 5.0 s), while strategy synthesis and orchestration logic consume the remainder. The workflow achieved a 100% success rate across all trials.

Crucially, the system's operational cost structure was highly predictable. LLMs process text in discrete units called tokens, roughly comparable to syllable-level fragments of a word, and the volume consumed per debate run averaged 4539 with negligible variance (σ = 25, CV = 0.54%). This 0.54% fluctuation is 27$\times$ lower than the temporal variance, indicating that the per-scenario computational demand is highly repeatable across all tested price conditions.

To translate this into economic terms, applying current pricing for efficient reasoning models (e.g., $0.15/1 M input tokens), the computational cost is approximately $0.0007 per

scenario. This renders the cost of multi-agent assessment negligible when compared to the potential NPV loss associated with a sub-optimized strategic decision.

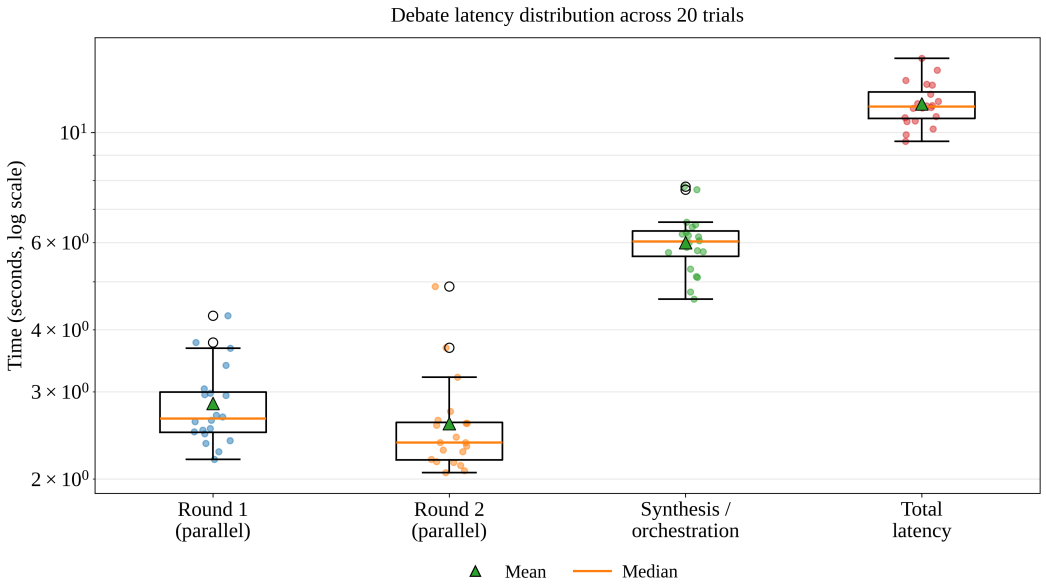


Figure 8. Debate workflow time distribution showing parallel agent execution and synthesis/orchestration overhead.

Table 4. System performance metrics across 20 controlled trials.

Metric	Mean	σ	Range
Debate Latency (s)	10.69	1.52	8.32–14.31
Processing Load (Tokens)	4,539	25	4494–4579
Est. Cost per Run (USD) *	$0.0007	<$0.0001	–

* Cost calculated using a standardized unit rate of $0.15 per million tokens.

Decision Coherence: Decision quality is assessed by examining whether the debate mechanism systematically adjusts user-proposed prices in a direction consistent with market context rather than accepting them uncritically. This constitutes a measurable and directionally rational behavior; whether it translates into superior planning outcomes in terms of pit value, production schedule quality, or other feasibility metrics remains an open question for future research. The trials also showed the economic reasoning capabilities embedded in the multi-agent debate structure. Table 5 reveals that the agents did not merely accept user proposals, but actively contextualized them against available market intelligence. When presented with conservative scenarios ($5.00/lb, 18% below spot), the system identified significant upside potential relative to the $6.07/lb spot price and prevailing bullish sentiment, recommending a consistent +15.0% adjustment to capture value that would otherwise remain stranded. For moderate scenarios ($5.80/lb), the system recommended modest increases (+2.59% to +4.31%) to better align with market conditions. Conversely, when users proposed optimistic scenarios ($6.50/lb, 7% above spot), the agents advised caution with corrections from −3.85% to −6.15% to buffer against peak-market volatility. This pattern confirms that the orchestration layer functions as a strategic governance step, ensuring that subjective engineering assumptions undergo stress-testing against external market realities before triggering resource-intensive optimization.

Table 5. AI economic behavior demonstrating strategic contextualization over rule-based correction.

Scenario (Input Price)	Mean Adj.	σ	Range
Conservative (\$5.00/lb, $n = 7$)	+15.00%	0.00	+15.00%
Moderate (\$5.80/lb, $n = 7$)	+3.51%	0.80	+2.59% to +4.31%
Optimistic (\$6.50/lb, $n = 6$)	−4.36%	1.02	−6.15% to −3.85%

5. Discussion

5.1. Architectural Positioning

MINDS is a transparent framework designed to integrate AI into the SMP toolchain. Established commercial platforms such as Vulcan [38], Surpac [39], and Whittle [40] provide mature and comprehensive capabilities for pit optimization, pushback design, and production scheduling, and remain the industry standard for detailed engineering work. Rather than competing with these tools, MINDS addresses the decision-support gap that precedes detailed design: the process by which engineers formulate, stress-test, and govern the economic assumptions that drive those downstream tools. By integrating conversational interfaces, multi-agent economic reasoning, and market intelligence into a governed workflow, the framework operates as an upstream complement to existing software.

Crucially, the framework separates architectural roles from specific intelligence implementations. By isolating interface, orchestration, and optimization layers behind stable database contracts, the design supports technology-agnostic substitution. While this study deploys multi-agent debate due to its effectiveness in structured critique [26,27], the orchestration backbone relies on interface standards rather than specific AI techniques. This avoids the 'lock-in' typical of rigid legacy systems, ensuring the SMP workflow can adopt new technologies incrementally.

For DMP specifically, MINDS addresses the workflow fragmentation that undermines scenario reproducibility. Traditional practice relies on manual parameter coordination across disconnected tools, where each iteration introduces version control risks. Table 6 positions the framework's capabilities relative to these conventional workflows, showing how automated orchestration and persistent audit trails eliminate the manual handoff patterns that constrain timely re-evaluation.

Table 6. MINDS capabilities relative to traditional DMP workflows.

Feature/Capability	MINDS	Traditional Workflow
Natural Language Interface	✓	✗
Multi-Agent Economic Evaluation	✓	✗
Pit Optimization Integration	✓	✓
Real-Time Market Intelligence	✓	✗
Human-in-the-Loop	✓ (explicit gates)	✓ (manual)
End-to-End Audit Traceability	✓	✗ (partial)
Modular Component Design	✓	✗ (monolithic)

The framework's modular design also defines its extensibility boundaries. While the current implementation targets open-pit, single-commodity deposits, the interface abstraction that isolates orchestration logic from domain-specific components allows for adaptation to other contexts. Multi-commodity deposits would require additional valuation logic in the Economic Assessment Layer to handle correlated price assumptions, while underground mining systems would necessitate substituting the optimization engine with solvers appropriate to stope sequencing or development scheduling.

5.2. Workflow Orchestration and Traceability

The framework automates data transfer between specialized components through database orchestration. Conversational interfaces, market data APIs, and optimization solvers interact through database contracts rather than spreadsheet handoffs, removing a chronic source of parameter misalignment. This matters particularly in DMP contexts, where market volatility demands frequent re-evaluation, yet traditional workflows accumulate friction with each iteration.

The architecture prioritizes persistent outcomes over transient outputs. Unlike conventional workflows that rely on unstructured file storage and manual forensic investigation to reconstruct decisions, MINDS transforms optimization results into queryable decision records. User intent, debate rationale, and market snapshots are stored as structured JSON records (i.e., a human-readable data format using named fields) and linked across scenarios using foreign keys (i.e., record-to-record links), enabling near-instant retrieval of the economic assumptions and debate rationale behind each scenario via SQL queries (i.e., database searches), making every scenario traceable to the assumptions and justification behind the decision.

5.3. Limitations

As a proof-of-concept, this study acknowledges three specific constraints that must be addressed before deployment in an industrial production environment:

- **AI Probabilistic Nature:** LLMs are not fully deterministic. While our trials showed high stability, industrial deployment requires strict guardrails to ensure the AI's commands remain consistent.
- **Data Complexity and Quality:** Demonstration relied on the synthetic *Marvin* benchmark and curated news feeds. Real-world deployment involves significantly higher complexity and requires high-quality inputs to ensure valid economic assessments.
- **Scope of Multi Agent Debate:** The current system evaluates commodity price. To support a full feasibility study, the debate logic must be expanded to assessing other critical drivers, such as operational strategies. The same architectural pattern could be extended to additional feasibility drivers, including production strategy, capacity constraints, CAPEX scenarios, and regulatory considerations, by introducing domain-specific agents and data feeds within the same orchestration backbone.

6. Conclusions

This paper introduced MINDS, a modular framework addressing integration challenges in SMP through orchestrated workflows spanning conversational interfaces, multi-agent assessment, and automated execution. Rather than building monolithic systems, the architecture separates functional roles from implementations, allowing for organizations to adopt specialized tools while maintaining integration through stable database contracts. Coupling real-time market intelligence with parametric models delivers analytical decision support without sacrificing the audit trails or engineer decision authority that corporate governance demands.

The three contributions introduced in Section 1 were evaluated in the Marvin case study. The modular reference architecture was validated by demonstrating that all four functional layers operate cohesively through technology-agnostic component binding, without requiring manual data transfer between planning stages. The conversational orchestration system was demonstrated end-to-end, from a single natural language request through pit optimization to delivered planning outputs. The multi-agent assessment module was evaluated using 20 controlled trials, confirming that the debate mechanism produces direc-

tionally rational price adjustments relative to market context across conservative, moderate, and optimistic price scenarios.

Testing with the Marvin benchmark established the framework's operational characteristics across three dimensions:

1. **Automated Orchestration:** The system coordinated conversational input, multi-agent review, and optimization stages through database contracts, executing complete workflows from natural language requests to delivered outcomes.
2. **Robust Stability:** Twenty controlled trials achieved 100% execution reliability with sub-11 s debate latency, demonstrating that orchestration remains stable despite stochastic LLM generation.
3. **Decision-making Traceability:** Database schema preserved decision history, linking optimization outcomes to economic assumptions while strictly enforcing approval gates that prevent autonomous execution.

The results show that workflow-level orchestration provides a practical path toward modernizing SMP infrastructure. By automating routine transitions while maintaining reproducible decision rationale, the framework enables timely scenario iteration in response to market volatility without abandoning the governance structures that prevent over-reliance on automated recommendations.

Future Directions

Three research directions merit attention: First, expert evaluation studies with practicing mine planners would assess the qualitative value of multi-agent debates and determine whether structured critique improves decision quality or introduces procedural overhead. Second, assessing the framework's architectural flexibility by substituting the multi-agent critique module with alternative assessment mechanisms. Although this work relies on a multi-agent debate approach, the framework's modular design allows this layer to be replaced by novel implementations as technology progresses without disrupting the broader orchestration backbone. Third, expanding the analytical scope beyond commodity prices to incorporate heterogeneous data streams, such as real-time operational signals, revised geological models, and regulatory changes, would determine whether the debate structure scales to the complexity of DMP. These extensions would advance AI-enhanced systems that maintain human oversight while addressing the fragmentation challenges that undermine current practice.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/mining6020026/s1.

Author Contributions: Conceptualization, R.N., N.R. and M.M.; Methodology, R.N.; Software, R.N.; Validation, R.N.; Formal analysis, R.N.; Investigation, R.N.; Resources, R.N.; Data curation, R.N.; Writing—original draft, R.N.; Writing—review & editing, R.N., N.R. and M.M.; Visualization, R.N.; Supervision, N.R. and M.M.; Project administration, R.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article and Supplementary Materials. Further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Accenture. The Mining Industry's $64 Billion Blind Spot. Technical Report, Accenture Research. 2025. Available online: https://www.accenture.com/us-en/blogs/mining-industry-64-billion-dollar-blind-spot (accessed on 3 February 2026).
2. International Energy Agency. *Global Critical Minerals Outlook 2025*; Technical Report; International Energy Agency: Paris, France, 2025. Available online: https://www.iea.org/reports/global-critical-minerals-outlook-2025 (accessed on 3 February 2026).
3. World Economic Forum. *The Future of Mining and Metals: Delivering the Energy Transition*; Technical Report; World Economic Forum: Geneva, Switzerland, 2025. Available online: https://www.weforum.org/stories/2024/06/energy-transition-critical-minerals-mining/ (accessed on 3 February 2026).
4. Sun, Y.; Ji, F. An Embodied Intelligence System for Coal Mine Safety Assessment Based on Multi-Level Large Language Models. *Sensors* **2025**, *25*, 488. [CrossRef]
5. Naqvi, S.M.R.; Ghufran, M.; Varnier, C.; Nicod, J.M.; Zerhouni, N. Enhancing semantic search using ontologies: A hybrid information retrieval approach for industrial text. *J. Ind. Inf. Integr.* **2025**, *45*, 100835. [CrossRef]
6. Wang, L.; Ma, C.; Feng, X.; Zhang, Z.; Yang, H.; Zhang, J.; Chen, Z.; Tang, J.; Chen, X.; Lin, Y.; et al. A Survey on Large Language Model Based Autonomous Agents. *Front. Comput. Sci.* **2024**, *18*, 186345. [CrossRef]
7. Guo, T.; Chen, X.; Wang, Y.; Chang, R.; Pei, S.; Chawla, N.V.; Wiest, O.; Zhang, X. Large Language Model based Multi-Agents: A Survey of Progress and Challenges. *arXiv* **2024**, arXiv:2402.01680. [CrossRef]
8. Wu, Q.; Bansal, G.; Zhang, J.; Wu, Y.; Li, B.; Zhu, E.; Jiang, L.; Zhang, X.; Zhang, S.; Liu, J.; et al. AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation. *arXiv* **2023**, arXiv:2308.08155. [CrossRef]
9. National Institute of Standards and Technology. *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*; Technical Report NIST AI 600-1; U.S. Department of Commerce, National Institute of Standards and Technology: Gaithersburg, MD, USA, 2024. [CrossRef]
10. ABB. ABB Launches AI-Powered GMD Copilot for Efficient Mining Operations. Mining Technology News. Available online: https://www.mining-technology.com/news/abb-gmd-copilot/ (accessed on 3 February 2026).
11. Airth. AI Powered Mining Solutions for the Future. Available online: https://www.airth.io/products (accessed on 3 February 2026).
12. Rojas, L.; Pena, A.; Garcia, J. AI-Driven Predictive Maintenance in Mining: A Systematic Literature Review on Fault Detection, Digital Twins, and Intelligent Asset Management. *Appl. Sci.* **2025**, *15*, 3337. [CrossRef]
13. Li, Y.; Wang, Z.; Liu, X.; Zhang, H.; Chen, J. MiningLLM: Towards Mining 5.0 via Large Language Models in Autonomous Driving and Smart Mining. *IEEE Trans. Intell. Veh.* **2024**, *9*, 3456–3469. [CrossRef]
14. Zhao, K.; Zhang, L.; Wang, X.; Liu, Y. Hierarchical Multi-Agent Reinforcement Learning for Sustainable Truck Dispatching in Open-Pit Mining. *Transp. A Transp. Sci.* **2025**, *21*, 456–478. [CrossRef]
15. Espinoza, D.; Goycoolea, M.; Moreno, E.; Newman, A.N. MineLib: A Library of Open Pit Mining Problems. *Ann. Oper. Res.* **2012**, *206*, 91–114. [CrossRef]
16. Noriega, R.; Pourrahimian, Y. A systematic review of artificial intelligence and data-driven approaches in strategic open-pit mine planning. *Resour. Policy* **2022**, *77*, 102727. [CrossRef]
17. Yang, N.; Liu, Z.; Zhao, X.; Xiang, J.; Zhang, R.; Liu, C.; Huang, Z.; Hu, G.; Zhao, Z.; Yuan, X.; et al. Artificial Intelligence for Mineral Exploration: A Review and Perspectives on Future Directions from Data Science. *Earth-Sci. Rev.* **2024**, *258*, 104941. [CrossRef]
18. Qi, C.; Fourie, A. Big Data Management in the Mining Industry. *Int. J. Miner. Metall. Mater.* **2020**, *27*, 131–139. [CrossRef]
19. Bakkula, O.; Wibisono, D.; Basri, M.H.; Siallagan, M. Towards resilience in the value chain of coal mining upstream: An agent-based modeling and simulation to improve coal discrepancy. *Discov. Appl. Sci.* **2024**, *7*, 22. [CrossRef]
20. Qian, Z.; Shi, C. Large language model-empowered paradigm for automated geotechnical site planning and geological characterization. *Autom. Constr.* **2025**, *173*, 106103. [CrossRef]
21. Wang, Y.; Feng, Y.; Xi, C.; Wang, B.; Tang, B.; Geng, Y. Development of an Intelligent Coal Production and Operation Platform Based on a Real-Time Data Warehouse and AI Model. *Energies* **2024**, *17*, 5205. [CrossRef]
22. Miñón, R.; López-de Armentia, J.; Bonilla, L.; Brazaola, A.; Laña, I.; Palacios, M.C.; Mueller, S.; Blaszczak, M.; Zeiner, H.; Tschuden, J.; et al. A multi-level IIOT platform for boosting mines digitalization. *Future Gener. Comput. Syst.* **2025**, *163*, 107501. [CrossRef]
23. Icarte-Ahumada, G.; Herzog, O. Intelligent Scheduling in Open-Pit Mining: A Multi-Agent System with Reinforcement Learning. *Machines* **2025**, *13*, 350. [CrossRef]
24. Sheikh, A.A.; Chong, E.K.P. Multi-Agent Reinforcement Learning Framework for Optimizing Smart Cities as System of Systems. *Syst. Eng.* **2026,** *29*, 3–19. [CrossRef]
25. Lin, J.; Zhang, W.; Wang, Q.; Liu, X. A New Simulated Annealing Algorithm for Simultaneous Open-Pit and Waste Dump Scheduling in Mining Complexes. *PLoS ONE* **2025**, *20*, e0333606. [CrossRef]
26. Du, Y.; Li, S.; Torralba, A.; Tenenbaum, J.B.; Mordatch, I. Improving Factuality and Reasoning in Language Models through Multiagent Debate. *arXiv* **2023**, arXiv:2305.14325. [CrossRef]

27. Liang, T.; He, Z.; Jiao, W.; Wang, X.; Wang, Y.; Wang, R.; Yang, Y.; Tu, Z.; Shi, S. Encouraging Divergent Thinking in Large Language Models through Multi-Agent Debate. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2024. [CrossRef]
28. Luo, Y.; Feng, Y.; Xu, J.; Tasca, P.; Liu, Y. LLM-Powered Multi-Agent System for Automated Crypto Portfolio Management. *arXiv* **2025**, arXiv:2412.20138. [CrossRef]
29. Rahman, M.M.; Terano, T.; Nishino, H. Multiagent Actor-Critic Generative AI for Query Resolution and Analysis. *Expert Syst. Appl.* **2025**, *238*, 122145. [CrossRef]
30. Jung, H.S.; Lee, H. Explainable zero-shot trading using multi-agent LLM architecture: A backtested approach for Bitcoin price. *Inf. Process. Manag.* **2026**, *63*, 104466. [CrossRef]
31. Kumral, M.; Dowd, P.A. A Simulated Annealing Approach to Mine Production Scheduling. *J. Oper. Res. Soc.* **2005**, *56*, 922–930. [CrossRef]
32. Ghali, M.K.; Pang, C.; Molina, O.; Gershenson-Garcia, C.; Won, D. Forecasting Commodity Price Shocks Using Temporal and Semantic Fusion of Prices Signals and Agentic Generative AI Extracted Economic News. *arXiv* **2025**, arXiv:2508.06497. [CrossRef]
33. Sun, Y.; Liu, X. Research and Application of a Multi-Agent-Based Intelligent Mine Gas State Decision-Making System. *Appl. Sci.* **2025**, *15*, 968. [CrossRef]
34. Sun, C.; Huang, S.; Pompili, D. LLM-Based Multi-Agent Decision-Making: Challenges and Future Directions. *IEEE Robot. Autom. Lett.* **2025**, *10*, 5681–5688. [CrossRef]
35. Lerchs, H.; Grossmann, I.F. Optimum Design of Open-Pit Mines. *Trans. Can. Inst. Min. Metall.* **1965**, *68*, 17–24. [CrossRef]
36. Hasozdemir, K.; Erçelebi, S.G. Enhancing the performance of integer models for addressing the long-term production planning problem in open pit mines by decision variable fixation based on parametric analysis of the final pit limit. *Gorn. Nauk. Tehnol. (Min. Sci. Technol.)* **2024**, *9*, 74–84. [CrossRef]
37. Alipour, A.; Khodaiari, A.A.; Jafari, A.; Tavakkoli-Moghaddam, R. An integrated approach to open-pit mines production scheduling. *Resour. Policy* **2022**, *75*, 102459. [CrossRef]
38. Maptek Pty Ltd. Vulcan: 3D Mine Planning and Geological Modelling Software. 2024. Available online: https://www.maptek.com/products/vulcan/ (accessed on 3 February 2026).
39. Dassault Systèmes GEOVIA. GEOVIA Surpac: Mine Planning and Geological Modelling Software. 2024. Available online: https://www.3ds.com/products/geovia/surpac (accessed on 3 February 2026).
40. Dassault Systèmes GEOVIA. GEOVIA Whittle: Strategic Mine Planning Software. 2024. Available online: https://www.3ds.com/products/geovia/whittle (accessed on 3 February 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.