Elevating safety and efficiency in mining with Vision AI: From object detection to large language model-driven decision intelligence

by Y. Zhang[1], M.A.H. Zahid[1], T. Moodley[2]

Affiliation:
[1]Hatch, Canada
[2]Hatch, South Africa

Correspondence to:
Y. Zhang

Email:
yale.zhang@hatch.com

Dates:
Received: 17 Oct. 2025
Published: February 2026

How to cite:
Zhang, Y., Zahid, M.A.H., Moodley, T. 2026. Elevating safety and efficiency in mining with Vision AI: From object detection to large language model-driven decision intelligence. *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 126, no. 2, pp. 135–140

DOI ID:
https://doi.org/10.17159/2411-9717/949/2026

ORCiD:
Y. Zhang
http://orcid.org/0009-0004-0735-3551
M.A.H. Zahid
http://orcid.org/0009-0003-3725-9825
T. Moodley
http://orcid.org/0009-0001-5908-7740

This paper is based on a presentation given at the 9TH International PGM Conference 2025, 27-28 October 2025, Sun City, Rustenburg, South Africa

Abstract

Mining operations are under growing pressure to improve safety and efficiency while dealing with aging infrastructure, complex processes, and workforce constraints. Although many sites are equipped with surveillance cameras and control systems, critical events often go unmonitored or under-analysed due to the lack of intelligent interpretation tools. Cameras typically act as passive recorders, requiring manual review by control room operators; a process that is labour-intensive, error-prone, and reactive. Vision AI is emerging as a transformative solution, combining computer vision and artificial intelligence to deliver real-time, actionable insights. This technology has evolved along two key phases: traditional object detection, and more recently, multimodal large language model integration. This paper presents solution architectures, deployment results, and key insights from real-world implementations across underground operations, open-pit truck-shovel operations, and smelter operations, demonstrating how Vision AI is reshaping mining operations to become safer, more efficient, and more intelligent.

Keywords

Vision AI, computer vision, large language models, mining safety, operational efficiency, object detection

Introduction

Mining operations face intensifying pressure to improve safety and productivity under both structural and operational constraints due to aging assets and infrastructure, increasingly complex processes, tighter environmental and social expectations, and a shrinking pool of experienced operators. Safety remains a top priority, yet persistent risks continue to threaten people, assets, and the environment. This indicates that the industry remains short of its vision of zero harm. On the productivity side, efficiency is crucial to remaining profitable, but there are limited ways to track performance, particularly for certain mining events that slip past traditional sensors.

To address these challenges, modern mining sites are equipped with extensive surveillance infrastructure and cameras, however this wealth of visual data remain largely underutilised. These feeds are often monitored manually in control rooms, turning video into a reactive, labour-intensive record rather than a continuous source of actionable intelligence. The human-centric monitoring approach also introduces variability in detection accuracy and response times across different operators and shifts. The result is a gap between what is visible and what is acted upon.

Vision AI, that is, computer vision enhanced by modern machine learning and artificial intelligence, addresses this gap by placing a smart engine behind each camera video stream to convert pixels into real-time, explainable signals. By automatically detecting critical safety issues, identifying operational bottlenecks, and offering meaningful recommendations, often faster and more accurately than manual processes, Vision AI solutions deliver substantial value to mining operations and are applicable across the entire value chain. Representative applications include:

- *Conveyor operations monitoring:* Detect overloading, spillage, and early indicators of belt tears or fire and issue timely alerts so that operators can intervene before minor anomalies escalate into production losses or safety incidents.
- *Primary and secondary crusher oversight:* Quantify truck cycle times, characterise ore size distribution, and assess feed conditions and fuse visual cues with operational data to expose inefficiencies and early signs of jams or overloads, reducing unplanned downtime.

- *Maintenance-bay safety and efficiency*: Track service activities on heavy equipment by detecting the suspended-load hazards, confined-space entry, and procedural deviations to provide supervisors with real-time visibility to protect personnel and shorten turnaround.
- *Rail and port operations:* Monitor rail crossings, loading compliance, and unauthorised intrusions to prevent accidents, reduce dwell time, and improve throughput across the transport interface. During marine loading, detect spillage and indicators of potential water contamination to trigger rapid responses while preserving evidence for regulatory compliance.
- *Dust monitoring and control:* Continuously estimate dust intensity and dispersion patterns from video and recommend adjustments to water-spray systems or operating parameters to mitigate health, environmental, and compliance risks.

This paper examines how Vision AI transforms mining operations by improving operational safety and efficiency. The discussion begins with an overview of technology evolution, highlighting the shift from traditional object-detection approaches to sophisticated multimodal systems that can understand operational context and provide insights using natural language. A practical Vision AI solution architecture is then presented, followed by two industrial case studies on Vision AI applications for open-pit load-and-haul operations and metal smelting operations. The paper concludes with key findings and future directions for Vision AI deployment in industrial environments.

Vision AI technology evolution

The first wave of industrial Vision AI was driven by fast, single-stage object detectors, most prominently the YOLO ("You Only Look Once") family, which localise people, equipment, and vehicles in video streams at real-time frame rates (Redmon et al., 2016; Bochkovskiy et al., 2020). In structured environments with well-labelled data, these models offer a strong capability to make speed vs. accuracy trade-off and have been widely studied and deployed for industrial operations and safety monitoring (Pengfei, 2022; Jonas et al., 2025). Their outputs readily identify activity, enable basic understanding, and populate dashboards for situational awareness.

A successful implementation by Hatch illustrates the potential of this approach. In an underground mine, the cage, essentially an elevator running along a vertical shaft, is used to transport materials, mobile equipment, and personnel between surface and underground levels. It often becomes an operational bottleneck, where any delays in moving the right materials or teams to the right place at the right time can cause major disruptions. In fact, production setbacks and budget overruns can increase as high as 30%. For this reason, it is crucial to monitor cage utilisation in real time and quickly identify any inefficiencies or performance gaps. To achieve this, shaft stations were equipped with cameras, and YOLO-based object detection models were developed and used to timestamp various events such as material runs, gas checks, personnel movements, shaft inspections, etc. Detected events were aggregated into structured logs and aligned with planned schedules, then visualised to expose deviations and bottlenecks in near real time. By continuously tracking cage utilisation and comparing actuals against plan, the Vision AI system can spot deviations right away before they snowball into bigger problems and therefore improve decision making on shift coordination and sequencing. At one client site, adherence to the production plan for muck and personnel movement improved by approximately threefold after the deployment of the Vision AI solution, with additional benefits from auditable safety records and data-driven schedule optimisation. Given that cage availability is a recurring bottleneck, delays can propagate to materially significant production and cost impacts. Managing cage efficiency, in the short term, improves day-to-day efficiency and better productivity among the workforce, including contractors. And in the long run, it leads to improved asset utilisation, enhanced overall productivity, and significantly fewer cost overruns.

Despite wide application across various industries, traditional object detection systems have notable limitations. They require large, labelled datasets specific to each deployment environment, must be retrained when camera angles or lighting conditions change, and lack the ability to understand context or relate visual information to procedural knowledge. They also struggle to detect unfamiliar or evolving events that fall outside their training parameters. These factors limit scalability and constrain the ability to deliver procedure-aware decision support.

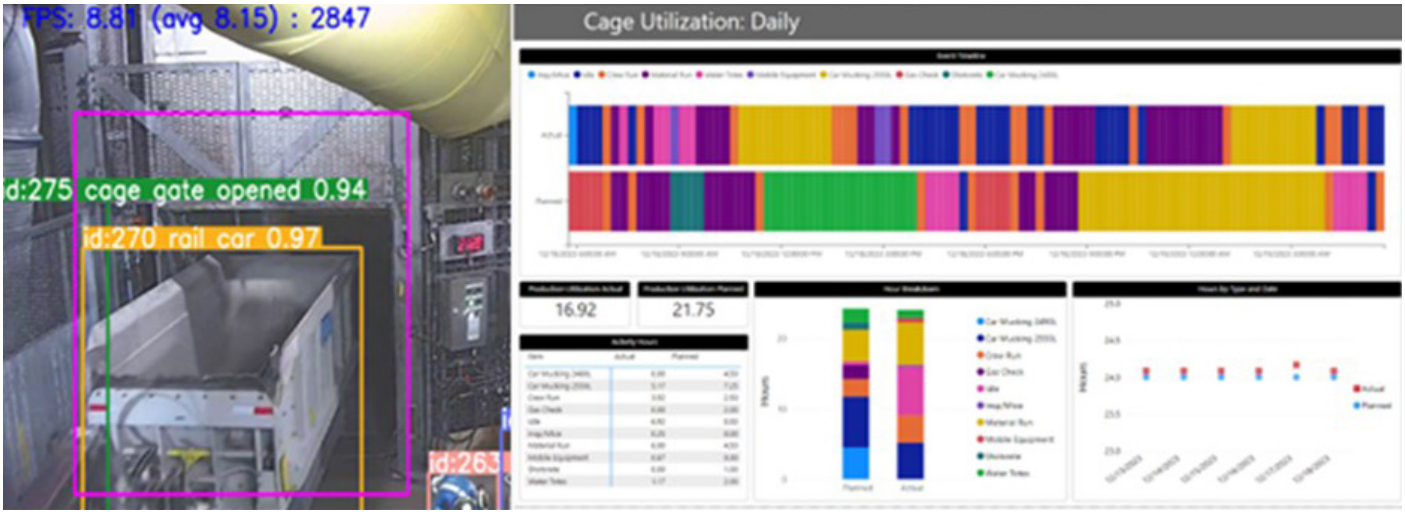

Figure 1—Example of underground mine cage utilisation dashboard powered by a YOLO-based Vision AI solution

Recent advances in generative AI and multimodal large language models (LLM) change this trajectory (Vaswani et al., 2017; Brown et al., 2020; OpenAI, 2023). Pretrained on diverse visual–text corpora, multimodal LLMs jointly process images/video and language, enabling systems that interpret event sequences, retrieve and apply site knowledge (SOPs, OEM manuals, etc.), and produce concise, human-readable notifications or summaries that explain why an alert is raised and what action should be taken. LLMs' instruction-following and few-shot conditioning ability reduces task-specific labelling demands of conventional computer vision and improves tolerance to camera drift and environmental variation. In effect, Vision AI evolves from frame-level detection to procedure-aware decision intelligence. It not only recognises events but also assesses their operational context and recommends next steps while preserving human oversight and traceability.

LLM-based Vision AI solution architecture

Figure 2 depicts the end-to-end Vision AI solution architecture (Zhang et al. 2025). A network of fixed or mobile cameras acquires continuous video in real-time across the industrial site. The video streams are time-synchronised and ingested at the edge, where lightweight vision modules perform denoising, exposure normalisation, stabilisation, and motion- or scene-change filtering to discard non-informative segments. This prescreening reduces bandwidth and computes while improving further analysis accuracy. For time- or mission-critical scenarios (e.g., restriction-zone breaches, conveyor fires), edge models execute real-time analyses and emit immediate alerts, even under intermittent connectivity. Store-and-forward buffers and watchdogs ensure continued operation during network outages.

Preprocessed clips, frames, and derived features are forwarded to a cloud reasoning layer that hosts multi-agent, multimodal large language models (LLM). In this layer, perception outputs (objects, trajectories, temporal segments) are fused with document database consisting of standard operating procedures (SOP), original equipment manuals (OEM), training materials, and with relevant telemetry e.g., SCADA tags, and dispatch events. Retrieval-augmented prompting grounds the models in site documents so that analyses are explainable and procedurally aware. Agent roles can be added for robustness, for example, an event assembly agent converts frame-level cues into state sequences; a procedure inspector agent tests conformance and assigns risk; and a recommendation agent generates operator-readable actions and structured outputs in a pre-configured JSON format that downstream systems can consume.

All edge- and cloud-generated artifacts such as event logs, embeddings, alerts, short video snippets, and model rationales are stored in a central data repository, which consists of a time-series/event store for analytics, object storage for media, and a vector index for fast document and scene retrieval. A role-based web user interface (UI) exposes real-time results, insights, and/or recommendations. In addition, it may also capture operator feedback (accept/override/annotate), which is fed back into continuous improvement pipelines. Optional integration can also be added to publish notifications to other systems such as process control systems, computerised maintenance management systems (CMMS), fleet management systems, or incident-management systems. This integration keeps humans in the loop while ensuring auditable traceability from alert to evidence to procedure.

The architectural choices target four operational properties.

- *Latency*: edge inference and event assembly keep the perception-to-alert path short for time- or mission-critical tasks, while less urgent reasoning (e.g., shift summaries) is batched in the cloud.
- *Reliability*: edge autonomy, health checks, and store-and-forward mitigate connectivity loss, models and configs are versioned and rolled out via a registry with canary/shadow modes.
- *Security and privacy*: video streams are redacted on edge (e.g., face/plate blurring) removing all PII from the streams, all data are encrypted in transit/at rest, and retention policies reflect regulatory and union requirements.

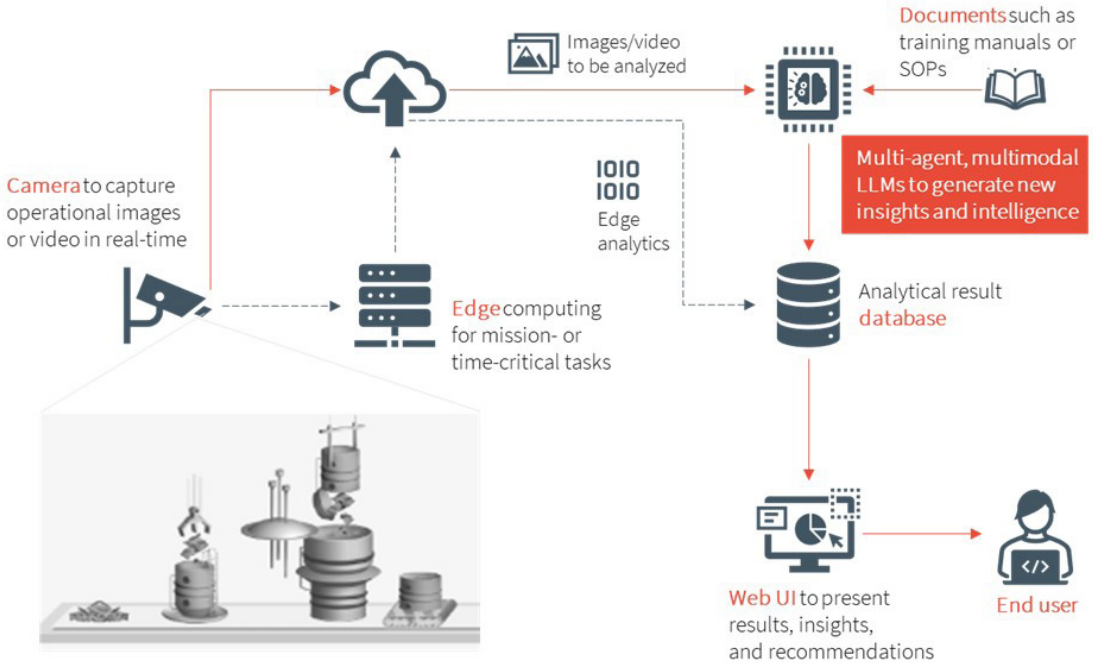

Figure 2—Vision AI solution architecture with embedded multimodal LLMs

- *Scalability*: stateless microservices and per-stream autoscaling allow sites to add cameras without redesigning.

Embedding LLMs into Vision AI solution provides three practical advantages.

- *First:* Reduced annotation burden – instruction-tuned, multimodal models adapt to new scenes with minimal task-specific labels, accelerating progression from pilot to production.
- *Second*: Contextual reasoning – models consider temporal order and retrieve site documents, turning "what is happening" into "why it matters" and "what to do next."
- *Third*: Robustness to variability – because reasoning is grounded in procedures rather than solely in pixel patterns, moderate changes in viewpoint, lighting, or equipment often require prompt updates rather than full retraining.

This architecture marries low-latency edge perception with document-grounded cloud reasoning. Cameras become intelligent sensors of which their outputs are not only visibility detections but contextualised reasoning and procedure-aware decisions.

Industrial case studies

Case Study 1: Open-pit mine operations

We applied multimodal large language models (LLM) to existing CCTV streams in an open-pit operation to derive actionable insight on productivity and safety without adding new instrumentation, as illustrated in Figure 3. The objective was twofold: (i) generate high-fidelity cycle analytics for trucks and shovels to support throughput improvement, and (ii) detect unsafe or inefficient behaviours early enough to enable proactive intervention.

Video feeds are ingested at the edge for basic stabilisation and motion filtering, then summarised events are passed to an LLM-enhanced reasoning layer. Perception outputs (e.g., equipment IDs, activity segments) are assembled into temporal sequences and interpreted against operating policies and safety guidelines. This produces operator-readable notifications and structured records suitable for dispatch and reporting systems.

The system reveals three types of information:

- Cycle analytics. Automatic identification of truck and shovel IDs; detection and timestamping of arrivals, queue/spot time, loading start/stop, travel, and dump events; computation of cycle-time distributions and dwell-time outliers by unit, bench, and shift.
- Operator performance assessment. Differentiation of efficient vs. inefficient behaviours (e.g., excessive spot-time variance, premature bucket withdrawal, repeated re-positioning) with evidence clips to support targeted coaching.
- Safety surveillance. Detection of large rock falls, overload and spillage, unsafe proximity/encroachment, and collision/near-miss precursors; graded alerts with concise rationales and links to the relevant procedure clauses when available.

Embedding LLMs in the pipeline reduces dependence on large, site-specific annotation campaigns. Instruction-tuned, multimodal models leverage few-shot exemplars and document grounding to adapt quickly to new pits, fleets, or revised safety protocols. Standard multimodal large language models primarily trained on general online datasets often lack the capability to deliver precise domain-specific analytics, such as cycle time measurements. To address this limitation, we fine-tuned a commercial multimodal model using domain-specific videos and analytics data. This approach resulted in an accuracy rate of approximately 98%. As a result, the solution tolerates moderate camera pose and illumination changes, shortens time-to-value, and lowers maintenance overhead while providing explainable outputs that can be audited and refined with operator feedback.

In operational terms, the approach shifts monitoring from retrospective review to continuous decision support: supervisors receive real-time alerts for emerging risks, dispatchers gain cycle-

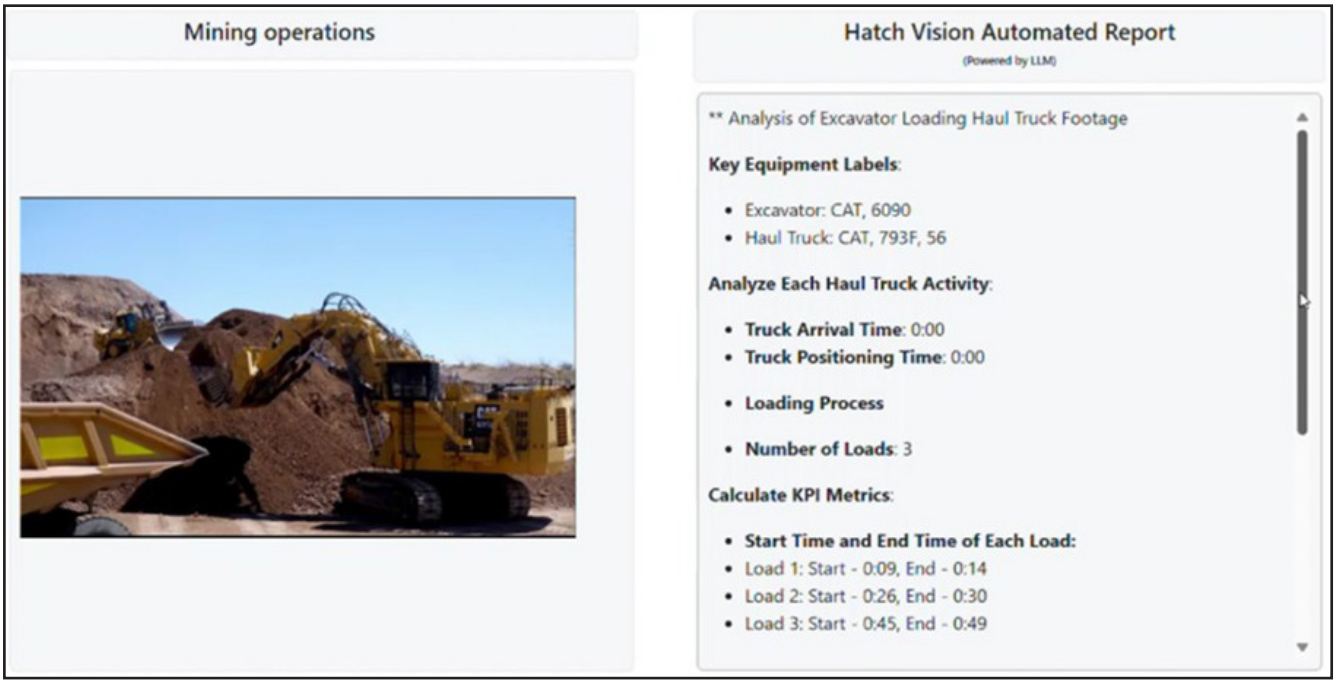

Figure 3—Vision AI platform to monitor open-pit load-and-haul operations

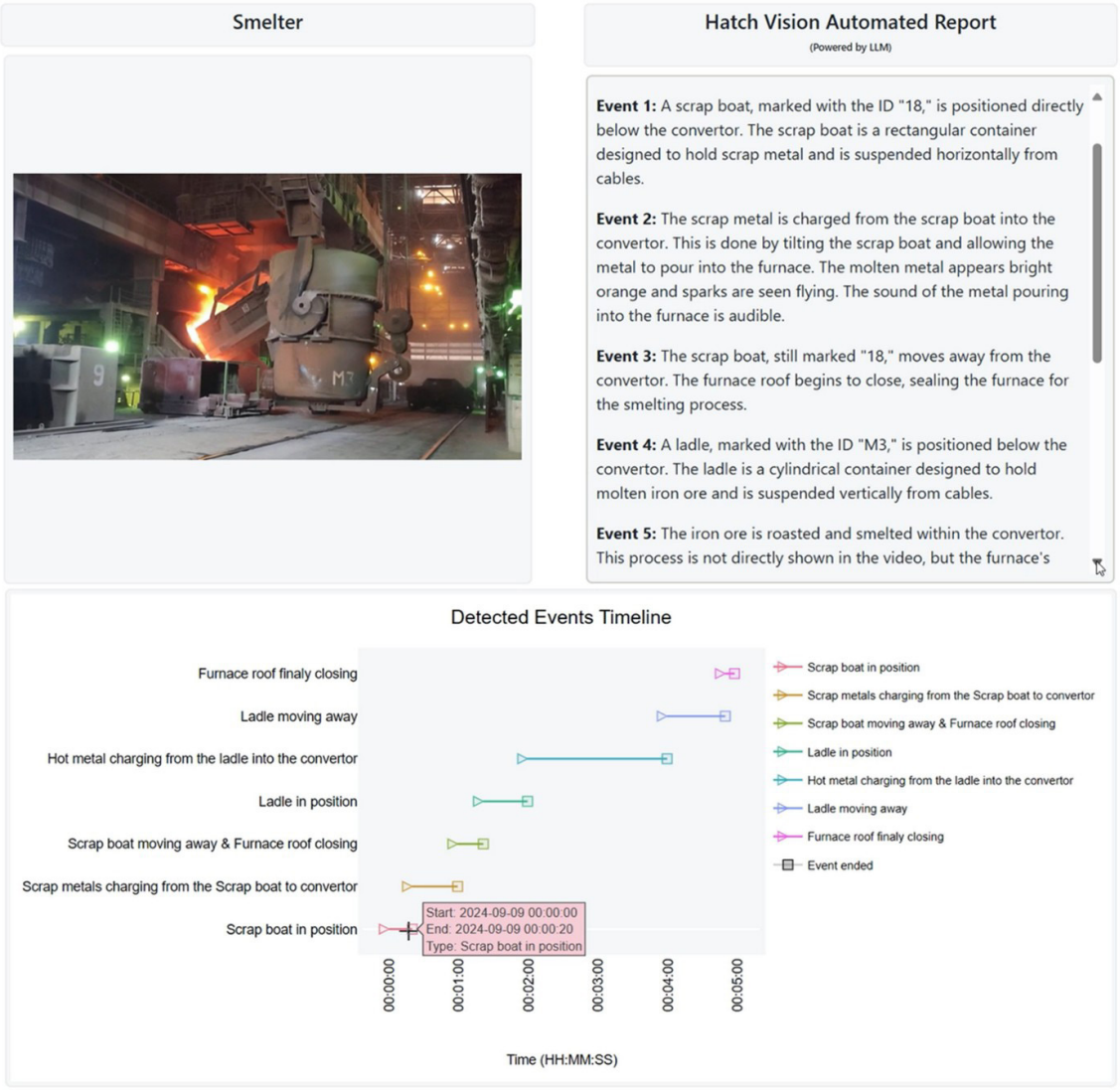

Figure 4—Vision AI platform identifying timing of key activities using LLM

time and queue insights to smooth flow, and training teams obtain objective evidence of behaviours to address. The net effect is faster issue resolution, reduced idle time and operational risk, and a clearer path to sustained throughput improvements.

Case Study 2: Smelter operations

This case study focuses on reliable identification of operational and safety events from smelter converter-aisle video, transforming unstructured footage into a time-stamped, evidence-linked record that supports logistics, scheduling, and risk control. A multimodal LLM reasons over assembled video sequences, retrieves relevant clauses from standard operating procedures (SOP), and emits both operator-readable explanations and structured event records.

Figure 4 illustrates the resulting artifacts. The upper panel shows a representative aisle frame with a synchronised, natural-language narrative of the detected sequence (e.g., scrap boat marked with ID '18' moves away, ladle marked with ID 'M3' is positioned below the converter). The lower panel presents a colour-coded Detected Events Timeline, where each bar denotes an event instance with its start and end timestamps (e.g., scrap boat in position, scrap metals charging to converter, ladle in position).

Applied to an industrial dataset spanning multiple shifts, the system can produce high-fidelity, time-stamped logs of key aisle activities and surface safety observations that may have been previously overlooked, for example, unauthorised intrusions or atypical charging behaviour. These outputs will be further analysed to improve operational safety and efficiency. Overall, the approach shifts smelter monitoring from retrospective review to continuous, procedure-aware decision support grounded in explicit event identification.

Practical consideration for Vision AI implementations

Performance comparison

Traditional object detection systems excel in specific, well-defined detection tasks where environmental conditions are stable and labelled training data is abundant. They can achieve high frame rates with low computational requirements, making them suitable for applications requiring immediate response to detected events.

Multimodal LLMs demonstrate superior contextual understanding and adaptation capabilities but require more computational resources for inference. They can interpret complex

visual scenes, understand relationships between objects, and relate observations to operational procedures without requiring extensive retraining for new scenarios.

Implementation considerations

Deploying Vision AI systems in mining environments requires careful consideration of hardware specifications and environmental protection. Cameras must withstand extreme temperatures, dust, moisture, and vibration while maintaining image quality sufficient for reliable analysis. Edge computing devices require industrial-grade specifications to operate reliably in harsh conditions.

Integration with existing mine management systems is crucial for maximising value. This includes data integration with historians, maintenance systems, and safety reporting platforms. Real-time processing requirements necessitate careful attention to system latency and reliability to ensure timely notifications while avoiding false positives.

Return on investment

The economic benefits of Vision AI implementation include direct cost savings from reduced labour requirements for monitoring, improved operational efficiency, and prevention of safety incidents. The case studies demonstrate measurable improvements: 23% increase in material movement and 31% reduction in scheduling deviations translate directly to increased capacity without additional capital investment.

Safety benefits, while more difficult to quantify directly, include reduced incident rates, improved compliance monitoring, and enhanced emergency response capabilities. The value of preventing a single serious safety incident often justifies the entire system investment.

Summary and conclusions

The evolution from traditional object detection to LLM-driven Vision AI represents a fundamental shift in monitoring capabilities for mining operations. While traditional approaches remain valuable for specific detection tasks, multimodal LLMs offer superior contextual understanding and adaptation capabilities that better match the complex, dynamic nature of mining operations.

The case studies demonstrate that Vision AI delivers significant improvements in both safety and operational efficiency across diverse mining applications. The shift from recognising what is happening to understanding why it matters represents the key advancement enabled by LLM integration, enabling more sophisticated operational insights and recommendations.

Future developments in multimodal AI technologies are expected to further enhance Vision AI capabilities, potentially enabling more autonomous operational monitoring and decision-making. Organisations that proactively develop Vision AI capabilities are likely to realise significant advantages in operational performance and safety outcomes.

References

Bochkovskiy, A., Wang, C.Y., Liao, H.Y.M. 2020. YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004. 10934.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S. 2020. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901.

Jonas, W., Hannes, B., Jan-Henrik, W., Shengjie, H., Tobias, H., Anas, A., Robert, S. 2025. Machine vision in manufacturing SMEs: a review. *Discover Applied Sciences*, vol. 7, pp. 371.

OpenAI. 2023. GPT-4 technical report. arXiv preprint arXiv:2303.08774.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A. 2016. You only look once: unified, real-time object detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788.

Ren, L., Wang H., Wang Y., Huang, K., Wang, L., Li, B. 2025. Foundation Models for the Process Industry: Challenges and Opportunities. *Science Direct Engineering*, vol. 52, pp. 53–59. https://www.sciencedirect.com/science/article/pii/S2095809925001766?via%3Dihub

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, A., Polosukhin, I. 2017. Attention is all you need. arXiv:1706.03762.

Xu, P., Zhou, Z., Geng, Z. 2022. Safety monitoring method of moving target in underground coal mine based on computer vision processing. *Nature Scientific Reports*, vol. 12. https://doi.org/10.1038/s41598-022-22564-8

Zhang, Y., Mohammed, Z., Marfatia, Z., Shah, A. 2025. Leveraging AI-Powered Large Language Models to Improve Operational Safety and Efficiency in the metal and Steel Industry. AISTech 2025 , Nashville, Tennessee USA ◆