Multi-Agent AI Framework for Objective Assessment of Excavator Operator Proficiency Through Temporal Behavior Analysis

Abstract

The construction industry faces a critical shortage of skilled heavy equipment operators, with traditional training assessment relying on subjective instructor evaluations that fail to capture the nuanced temporal coordination patterns distinguishing expert from novice performance. This study presents a novel multi-agent artificial intelligence framework for objective assessment of excavator operator proficiency through comprehensive temporal behavior analysis. The framework employs specialized AI agents to process synchronized data streams from industry-standard Caterpillar simulators, including joystick control signals, video recordings, and operational telemetry. Three mathematical formulations, the Pairwise Simultaneity Index (PSI), Bimanual Coordination Score (BCS), and N-Control Usage metrics, serve as the analytical foundation enabling quantitative assessment of operator coordination patterns and skill progression. Data were collected from 119 operators across three skill levels using standardized dig-footing exercises. Results revealed systematic differences in temporal coordination, with expert operators achieving significantly higher simultaneity indices (Arm-Boom: 0.542 vs 0.125 for novices) and bimanual coordination scores (0.333 vs 0.109). Multi-control usage analysis showed experts utilized three or more simultaneous controls for 35.7% of operational time compared to 6.3% for novices, indicating qualitative shifts in motor control organization rather than linear skill accumulation. The AI framework successfully generated personalized training reports that diagnosed specific coordination deficiencies and provided targeted intervention recommendations, demonstrating practical value for scalable operator development. This research contributes the first comprehensive framework integrating temporal behavior analysis with multi-agent AI for construction equipment training, providing objective skill assessment capabilities that serve as an assistive tool for instructors while enhancing evaluation precision and consistency.

Keywords: Construction equipment training, multi-agent AI, temporal behavior analysis, operator skill assessment, excavator simulation.

Authors: Sueed W., Muhammad K., Asif C., Chao W.

1.0 Introduction

The construction industry faces a critical shortage of skilled heavy equipment operators, with an estimated 430,000 unfilled positions across the United States as of 2024 (ABC, 2022; AGC, 2024). This skills gap directly impacts project timelines, operational efficiency, and safety standards on construction sites worldwide (Fulford & Standing, 2014). Excavator operation, in particular, requires sophisticated coordination skills that traditionally take years to develop through apprenticeship programs and on-the-job training (NCCER, 2023). The economic implications are substantial, with inadequately trained operators contributing to reduced productivity, increased fuel consumption, equipment wear, and elevated safety risks that can result in costly project delays and workplace accidents. Current operator training programs rely heavily on subjective assessment methods, where human evaluators make qualitative judgments about operator proficiency based on observation and experience. While industry standards such as those developed by the National Center for Construction Education and Research (NCCER) provide structured competency frameworks, the actual assessment process remains fundamentally subjective, creating inconsistencies in skill evaluation and certification outcomes (NCCER, 2023). These traditional approaches fail to capture the nuanced temporal dynamics of operator behavior, such as the coordination patterns and timing strategies that distinguish expert operators from novices (Bernold, 2007). Furthermore, existing assessment methods provide limited insight into specific skill deficiencies, making it difficult to design targeted training interventions that address individual operator needs.

The emergence of advanced multi-agent artificial intelligence systems presents opportunities to revolutionize operator skill assessment and training. Recent developments in agentic AI frameworks have demonstrated the capability to deploy complex, collaborative intelligence systems that can process multiple data streams simultaneously and generate sophisticated analytical outputs. These developments are enabling specialized agents work together in concert, contributing their own domain expertise without losing system-wide coordination (Fan et al., 2021a). These technological advances, combined with the widespread adoption of high-fidelity training simulators, create the foundation for objective, data-driven operator assessment systems that can provide real-time feedback and personalized training recommendations. The use of high-fidelity simulators that replicates the real equipment has become the industry standard for operator training, utilized by construction companies, training institutions, and military organizations globally. These simulators generate rich behavioral data streams including precise joystick control signals, video recordings of operator performance, and comprehensive telemetry data covering productivity metrics, fuel consumption, and operational timing (J. S. Lee et al., 2022; Sekizuka et al., 2020). Importantly, hundreds of performance benchmarks based on expert operator data are embedded within these systems, providing validated reference standards for skill assessment (Miyazaki et al., 2021). The simulator environment replicates authentic excavator dynamics with engineering-grade precision, ensuring that operator behavior captured during training sessions accurately reflects real-world operational patterns. Despite the wealth of data generated by modern training simulators, current systems lack analytical frameworks capable of transforming raw

behavioral signals into actionable insights about operator skill development. The temporal coordination patterns that characterize expert performance and timing optimization, remain largely unquantified (Dunston et al., 2014). This represents a significant opportunity to develop artificial intelligence systems that can decode these complex behavioral patterns and translate them into objective skill assessments that exceed human evaluator capabilities in both precision and consistency.

This research introduces a novel multi-agent AI processing framework specifically designed to analyze excavator operator performance through comprehensive temporal behavior analysis. The framework employs specialized AI agents working collaboratively to process video recordings, joystick control signals, and operational telemetry data, transforming these inputs into quantitative skill metrics and personalized training recommendations.The system generates comprehensive performance reports that provide objective skill assessments and targeted training interventions, addressing the fundamental limitation of subjective evaluation methods in current operator training programs. The primary contribution of this work is the development of an intelligent assessment system that leverages the same simulator platforms used for operator training across the construction industry. By utilizing data collected from Caterpillar's Advanced Construction Excavator Simulator System (CAT, 2025) during actual training sessions with operators ranging from novice to expert levels, this research demonstrates how multi-agent AI can objectively quantify operator skill and generate personalized feedback that accelerates skill development. The resulting framework provides a scalable solution for heavy equipment training schools and construction companies seeking to optimize training efficiency, reduce certification time, and ensure consistent operator competency standards across diverse training environments.

2.0 Literature Review

2.1 Operator Skill Assessment in Construction Equipment Training

The construction industry faces a fundamental challenge in developing skilled equipment operators capable of safely managing increasingly sophisticated machinery. Traditional assessment methods have relied heavily on subjective instructor evaluations and simple time-based metrics, creating inconsistencies in skill measurement and training effectiveness (cite). Bernold (2007) recognized this limitation and introduced quantitative assessment methods using electronic sensors to capture both operator control inputs and equipment responses. This pioneering work established that objective measurement could differentiate skill levels more reliably than subjective observation. Building on this foundation, Miyazaki et al., (2021) advanced the concept by developing database-driven skill evaluation systems for hydraulic excavator operations, demonstrating how systematic data collection could transform operator assessment from intuitive judgment to evidence-based evaluation. These developments indicate a trend in current research, moving from traditional apprenticeship-style assessment to scientifically grounded measurement approaches. The emergence of simulation technologies has further revolutionized operator assessment by providing controlled environments for skill evaluation without real-world safety risks. Sekizuka et al., (2020) demonstrated that remotely controlled excavators combined with virtual reality could reliably distinguish between expert and novice operators through multiple performance indices calculated from recorded operational data. Their validation studies revealed strong correlations between simulation-based metrics and real-world equipment performance, establishing simulation as a legitimate assessment platform. The integration of virtual reality has proven particularly valuable for crane operation training, where Song et al., (2021) found that VR-based systems significantly improved trainee self-efficacy across different crane types, with effectiveness mediated by user presence and perceived system usefulness. However, as Tichon & Phil, (2010) noted in their analysis of simulation benefits and limitations, these technologies require careful implementation strategies to maximize their training effectiveness.

Modern assessment approaches increasingly emphasize real-time feedback and instructor support systems to address the challenge of monitoring multiple trainees simultaneously. Vahdatikhaki et al., 2022) developed feedback support systems that capture trainee performance and automatically identify areas requiring instructor attention, enabling more objective communication between instructors and trainees during practical sessions. This approach recognizes that effective skill development requires timely, specific feedback that traditional observation methods struggle to provide consistently. Similarly, Ahn et al., (2020) demonstrated that BIM-enhanced training environments produced significantly better learning outcomes than conventional lecture methods, highlighting the importance of interactive, technology-supported learning experiences. Despite these technological advances, current assessment methods remain limited by their focus on individual performance metrics rather than comprehensive skill evaluation. Most systems still rely primarily on completion time as the primary assessment criterion, which proves insufficient for detailed skill evaluation and fails to capture the complex cognitive and motor coordination required

for expert operation. This limitation becomes particularly problematic when considering that expert operators often work more deliberately in complex situations, prioritizing safety and precision over speed Sekizuka et al., (2020). The field lacks integrated assessment frameworks capable of simultaneously evaluating multiple dimensions of operator competency, including technical skill, safety awareness, efficiency, and adaptability to varying operational conditions. This gap presents a significant opportunity for multi-agent AI systems that could synthesize diverse performance indicators into comprehensive skill profiles while providing personalized training recommendations tailored to individual operator development needs.

2.2 Control Signal and Behavioral Analytics in Equipment Operation

The analysis of control signals has become an important approach for understanding human performance in complex human–machine interaction systems. Control inputs generated through devices such as joysticks and levers represent the physical manifestation of operator decision-making, motor coordination, and task execution strategies (Bae et al., 2019). Early work on joystick-based human operations showed that point-to-point control movements consist of partially overlapping sub-movements with predictable velocity characteristics, demonstrating that joystick signals encode meaningful behavioral information beyond simple position commands (Slutski et al., 2000). This line of work established a foundation for interpreting control signals as measurable expressions of operator behavior, coordination, and skill development rather than merely machine input records. Advances in sensing technology and signal processing have enabled researchers to extract increasingly sophisticated behavioral indicators from joystick control data. In broader human–machine interaction domains, joystick signals have been used to examine operator cognitive states, workload, and performance characteristics in simulated operational environments. Liu et al., (2014) analyzed pilot turning behavior and cognitive load using simulated flight data, while Taheri Gorji et al., (2023) combined machine learning and physiological sensing to discriminate pilot cognitive workload during flight operations. Although these studies were conducted outside construction, they are important because they demonstrate that joystick control patterns contain distinctive temporal and behavioral signatures associated with expertise, workload, and task performance. Together, they support the broader premise that control-signal analysis can reveal underlying operator states relevant to training and assessment.

Within equipment operation research, signal-based analysis has become increasingly relevant because many machines require coordinated manipulation of multiple degrees of freedom. Construction equipment operation presents a particularly demanding context due to the need for bimanual control, simultaneous function blending, and adaptation to dynamic task conditions. In excavator operations specifically, tasks such as digging, trenching, leveling, swinging, and dumping involve coordinated movement of boom, stick, bucket, and swing functions through joystick inputs. Bae et al., (2019) showed that joystick signal measurements can be used to identify excavator working cycles using dynamic time-warping methods, demonstrating that earthmoving activities exhibit recognizable operational patterns even when temporal alignment differs across operators. Their work highlighted both the promise and the challenge of joystick-based activity

recognition, noting that signal characteristics vary with machine type, operator behavior, and skill level. Control mapping and interface design also influence the interpretation of joystick behavior and the transferability of learned skill. Proctor et al., (2022) compared Society of Automotive Engineers (SAE) and backhoe loader (BkL) control mapping patterns in excavator simulators and found that different control schemes significantly affect operator performance and skill transfer. This finding is particularly relevant for training research because it shows that control behavior is shaped not only by operator proficiency but also by interface structure. As a result, any signal-based assessment framework must account for how control mappings influence learning patterns, coordination strategies, and movement efficiency.

Beyond activity classification, understanding how operators coordinate multiple control axes simultaneously constitutes a distinct and underexplored dimension of skill analysis. This problem is fundamentally one of measuring co-activation across binary activation sequences, determining whether and how often two or more joystick axes are active at the same time. Quantifying co-activation across binary sequences has precedent in other domains. The Jaccard index, originally formulated as a means of characterizing general co-occurrence relations, is well-suited to the analysis of paired binary data where the central question is whether two binary characteristics co-occur more often than would be expected by chance (Fletcher & Islam, 2018; Luciano da F, 2021). Because the Jaccard index focuses only on mutual presences rather than mutual absences, it is particularly appropriate for asymmetric binary data where zero co-occurrence carries little informational weight, precisely the situation in joystick telemetry, where both axes being simultaneously inactive reflects operator pause rather than coordination (Vorontsov et al., 2013).

Equally important to pairwise co-activation is the question of how both hands work together across control axes. Rueda-Delgado et al., (2014) established that temporal and spatial parameters jointly constrain the coordination of limb movements, reflecting a basic synchronization tendency that can be shaped or suppressed through skill acquisition and practice. Behavioral studies have further demonstrated that the fundamentals of bimanual coordination are not a simple sum of unimanual activities, and that acquiring a novel bimanual coordination policy, synchronizing sequences of movements from both hands, represents a higher-order learning demand than unimanual motor skill development. In the context of heavy equipment operation, this cross-hand coordination is operationally critical: skilled excavator operators must blend left and right joystick inputs across multiple degrees of freedom simultaneously, and the degree to which they do so reliably distinguishes novice from expert performance (Bernold, 2007). More broadly, recent research has emphasized that control-signal analysis should move beyond isolated position or velocity measurements toward more integrated representations of operator behavior. Advanced signal-processing approaches increasingly focus on temporal relationships, coordination patterns, and multivariate interactions across control channels rather than treating each signal independently. This shift is important because skilled machine operation often depends less on the magnitude of individual inputs than on how multiple controls are synchronized over time. Behavioral signal processing research has similarly emphasized that human-generated signals are inherently variable

and context-dependent, reflecting complex interactions among operator characteristics, task demands, and environmental conditions (Narayanan & Georgiou, 2013). In construction equipment settings, this variability is amplified by differences in machine configuration, work scenario, and experience level, which complicates the development of robust skill indicators. Despite these advances, many existing behavioral analysis methods remain limited in their usefulness for training evaluation. A large portion of prior work has focused on activity recognition, workload inference, or general control characterization rather than developing interpretable coordination metrics that can directly support instructor decision-making. Even when signal-based indicators are extracted successfully, they are often not translated into actionable feedback for operator development. This creates a gap between analytical capability and training application. In practical equipment training, instructors need metrics that do more than describe control behavior; they must also identify meaningful deficiencies, distinguish skill levels, and guide targeted intervention.

2.3 Artificial Intelligence and Multimodal Systems for Training and Feedback

Translating quantitative coordination metrics into pedagogically meaningful feedback represents a distinct challenge that has motivated the integration of artificial intelligence into operator and motor-skill training research. Early efforts in equipment training applied conventional rule-based scoring and dashboard visualizations to instructor support, with Vahdatikhaki et al. (2022) demonstrating that automated highlighting of trainee performance enables more focused instructor intervention while still requiring expert interpretation of the underlying metrics. Parallel work in construction safety training has shown that immersive, technology-enriched learning environments produce measurable gains over conventional instruction (Ahn et al., 2020b). Outside construction, comparable progress has been made in medical motor-skill assessment, where deep learning models trained on operative video have demonstrated that expert and novice performance can be reliably discriminated from temporal features alone (Funke et al., 2019). Computer-vision-based deep learning has, in parallel, become an established tool for construction-site hazard recognition, personal protective equipment compliance, and worker activity monitoring, yielding accurate detection performance across diverse field conditions (J. Lee & Lee, 2023). Collectively, these strands establish that machine learning can convert raw behavioural and visual inputs into evaluative signals; however, prior systems have generally produced classifications, alerts, or numerical scores rather than the structured, narrative feedback that instructors deliver in apprenticeship-style training.

The emergence of foundation models and multimodal large language models (MLLMs) has begun to alter what is feasible at this evaluative-to-pedagogical interface. Foundation models, defined as large neural networks trained on broad data and adapted to a wide range of downstream tasks (Bommasani et al., 2022), exhibit task-agnostic generalization that earlier task-specific architectures could not achieve (Brown et al., 2020). Successive multimodal extensions, beginning with GPT-4 (OpenAI et al., 2024) and continuing through video-capable long-context models such as Gemini 1.5 (Team et al., 2024) accept interleaved text, image, and video inputs natively and

reason over extended temporal contexts spanning hours of footage in a single inference. This capability is particularly relevant for operator training applications, where assessment requires reasoning over the temporal evolution of behaviour rather than a single visual snapshot, and where heterogeneous evidence such as session video, joystick coordination plots, and tabular performance reports must be jointly considered. Retrieval-augmented generation has further demonstrated that grounding language-model outputs in domain-specific reference material reduces hallucination and improves factual reliability for knowledge-intensive tasks (Lewis et al., 2021), suggesting that benchmark coordination data and instructor-curated training rubrics can be leveraged as authoritative anchors during automated feedback generation. Taken together, these developments enable systems that not only recognise behavioural patterns but can also articulate them in coherent, instructor-facing language.

Deploying a single MLLM as a monolithic evaluator, however, conflates heterogeneous analytical tasks, including video segmentation, numerical reasoning over signal metrics, chart interpretation, and pedagogical synthesis, into one prompt and produces outputs that are difficult to inspect, attribute, or modify. Multi-agent AI architectures address this limitation by decomposing complex problems across specialised agents that retain their own roles, instructions, and reasoning traces while collaborating through structured message passing (Fan et al., (2021b); Xi et al., (2023)). Frameworks such as AutoGen (Wu et al., n.d.-a) formalise this decomposition by allowing developers to compose conversable agents around well-defined responsibilities and orchestration patterns, while Park et al. (2023) demonstrate that language-model-driven agents can sustain coherent, context-aware behaviour over extended interactions when equipped with explicit memory and reflection mechanisms. In assessment pipelines, this paradigm offers two practical benefits relevant to the present study. First, modularity permits independent substitution and evaluation of each analytical stage, supporting principled experimentation and reproducible benchmarking. Second, sequential context propagation preserves the narrative dependencies between stages, so that downstream synthesis agents can ground their judgments in the structured outputs of upstream specialists rather than re-deriving them from raw inputs.

Despite the maturity of these three threads, deep computer-vision systems for construction monitoring (J. Lee & Lee, 2023), generalist multimodal large language models ((OpenAI et al., 2024); (Team et al., 2024)), and multi-agent orchestration frameworks ((Wu et al., n.d.-b); (Xi et al., 2023)), no published work integrates them into an end-to-end pipeline that ingests operator joystick telemetry, simulator session video, and per-trial machine-generated performance reports and produces coordinated, multi-source operator feedback aligned with established skill-development rubrics. Existing construction-AI applications focus predominantly on safety surveillance and activity recognition rather than instructor-style coaching; existing tutoring and feedback systems for equipment operation do not yet exploit multimodal foundation models; and existing multi-agent demonstrations have been concentrated in software engineering, gameplay, and conversational domains rather than heavy-equipment skill assessment. The framework introduced in the methodology is positioned to close this gap, employing a sequential multi-agent

vision-language pipeline in which specialised Gemini-based agents handle video reasoning, telemetry analysis, chart interpretation, and pedagogical synthesis, and then assemble a self-contained operator training report that translates coordination metrics into actionable, evidence-grounded recommendations.

3.0 Methodology: AI-Agent-based Excavator Operation Feedback

3.1 System Overview

The proposed system is a multi-agent, vision-language pipeline implemented in Python and coordinated by a *ReportOrchestrator* module that governs the end-to-end report-generation workflow. As illustrated in Figure 3.1 and Figure 3.2, the pipeline accepts a video recording of a simulator training session (supplied as a URL), joystick telemetry data (in JSON format), and a simulator-generated performance report (in PDF format), and outputs a self-contained, AI-authored operator training report. Each processing stage is delegated to a purpose-built agent, thereby enabling modular substitution of individual components and facilitating systematic experimentation with alternative large language models (LLMs) or sampling strategies without requiring modification of the surrounding pipeline.

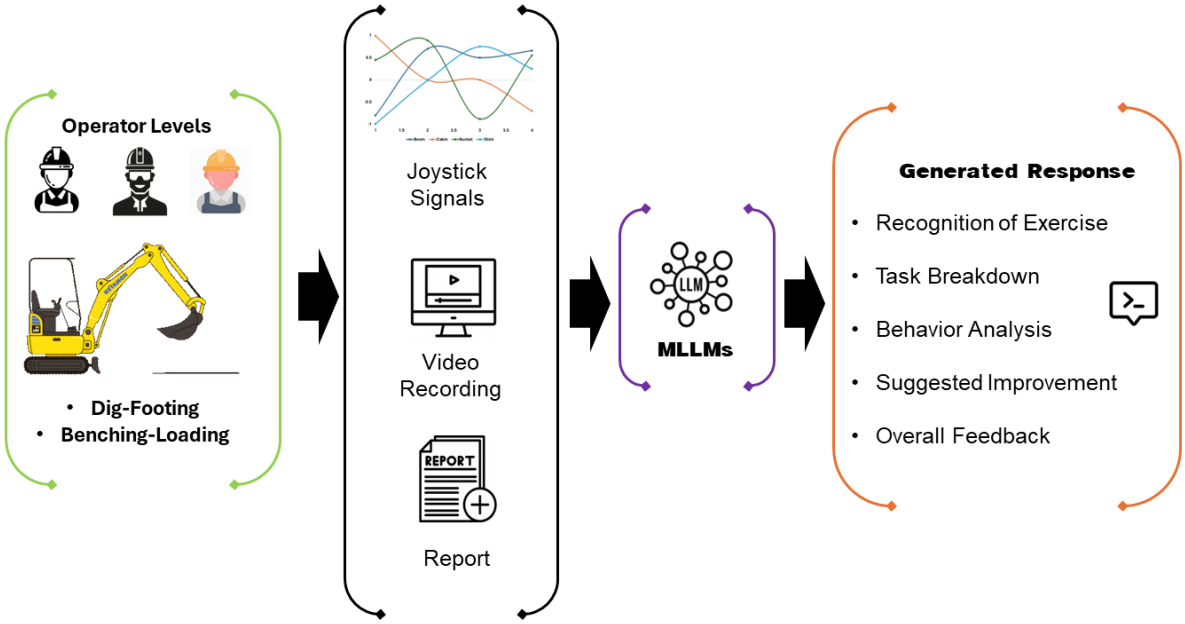


Figure 3-1. AI Agent-based Framework

The architecture follows a sequential, context-passing design in which the output of each agent constitutes part of the input context for the subsequent stage. This design was adopted in preference to parallel execution for two principal reasons. First, later stages such as insights generation depend on the combined outputs of earlier analytical stages, and sequential orchestration eliminates the need for complex synchronization barriers. Second, narrative coherence is enhanced when each agent can condition on the accumulated context produced by its predecessors, yielding a more logically consistent final report. Three Google Gemini model tiers are integrated within the pipeline: Gemini 2.5 Pro is employed for the primary video analysis stage, where deep temporal reasoning over extended video sequences is required; Gemini 3 Pro Preview serves as the multimodal large language model for the joystick-analytics, chart-analysis, and insights-generation stages; and Gemini 2.0 Flash is used for the template-driven report-assembly stage, where inference speed is prioritized over deep reasoning capacity.

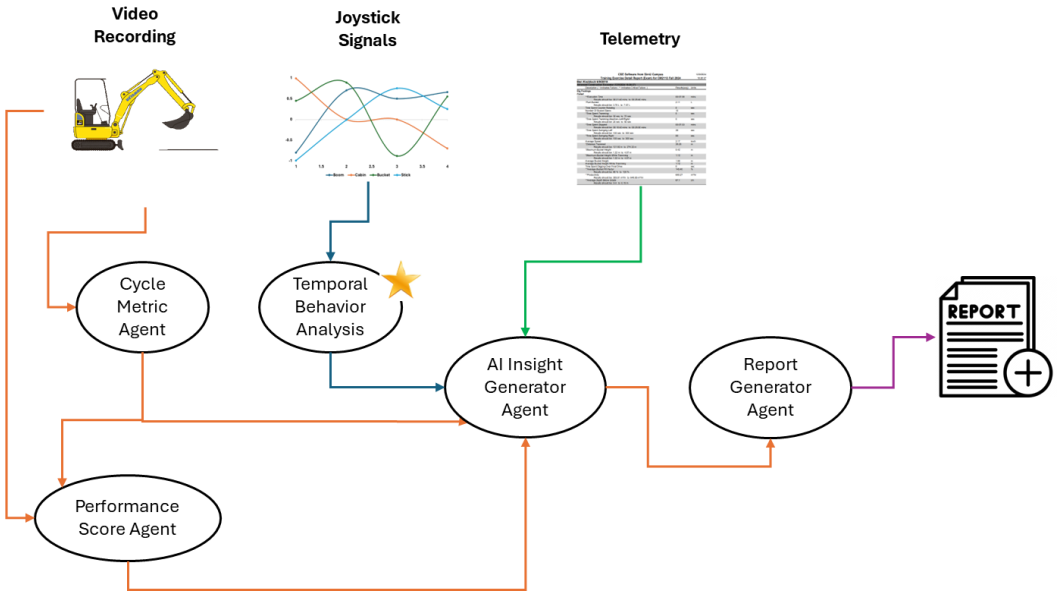


Figure 3-2. Multimodal Architecture

3.2 Excavator Simulator Setup and Training Protocol

The study utilized the CAT excavator simulator, a comprehensive training system featuring adaptable OEM controls, multi-monitor displays, and motion functionality. The simulator replicates the CAT 336D excavator's motion, vibration, and joystick controls with high fidelity. The system generates performance metrics including fill factor, productivity, fuel consumption, and cycle time after each exercise. Participants were categorized by experience level: novice (<1 year, n=78), intermediate (1-3 years, n=21), and expert (>3 years, n=20). The imbalance reflects the scarcity of expert operators in the construction industry. All participants completed the simulator's control familiarization program to ensure understanding of SAE-pattern joystick operations before assessment. The evaluation task was the dig-footing exercise, a standardized excavation involving digging a foundation area to specified depth (indicated by green marking) and depositing material in a designated dump zone (Figure 3). Training concluded when participants could complete the task independently without instructor guidance.

3.3 Data Sources

Three complementary data streams are collected for each training trial and subsequently fused within the pipeline. The integration of heterogeneous modalities such as visual, kinematic, and simulator-computed enables a more comprehensive assessment of operator performance than any single data source could provide in isolation.

Video Recordings: Simulator sessions are recorded as MP4 video files capturing the full operator interface, including on-screen control overlays and the three-dimensional simulation viewport. Video constitutes the primary observational modality, preserving the temporal dynamics of operator actions and enabling the precise temporal annotation of excavation cycles.

Joystick Telemetry: For each trial, the simulator exports a JSON file containing pre-computed joystick coordination metrics. These comprise: (1) a 4 × 4 Simultaneity Index (SI) matrix

quantifying the proportion of active time during which each pair of control axes is operated concurrently; (2) a scalar Bimanual Coordination Score (BCS) derived from the SI matrix; and (3) a simultaneous control-usage distribution reporting the fraction of active time during which one, two, three, or all four control axes are engaged. Companion visualization files, a Simultaneity Index heatmap and a simultaneous control-usage bar chart, accompany the JSON data to support downstream visual analysis. Joystick telemetry captures fine-grained motor-coordination patterns that are not readily discernible from video alone, thereby providing a quantitative basis for evaluating bimanual dexterity.

Simulator Performance Report: The excavator simulator automatically generates a per-trial PDF performance report containing machine-computed productivity metrics. The metrics extracted by the pipeline include material productivity (m^3/hr), fuel consumed during the session (L), and cumulative superstructure swing time to the left and right (seconds). These objective efficiency indicators complement the behavioral and coordination analyses by situating operator technique within the context of measurable task-level outcomes.

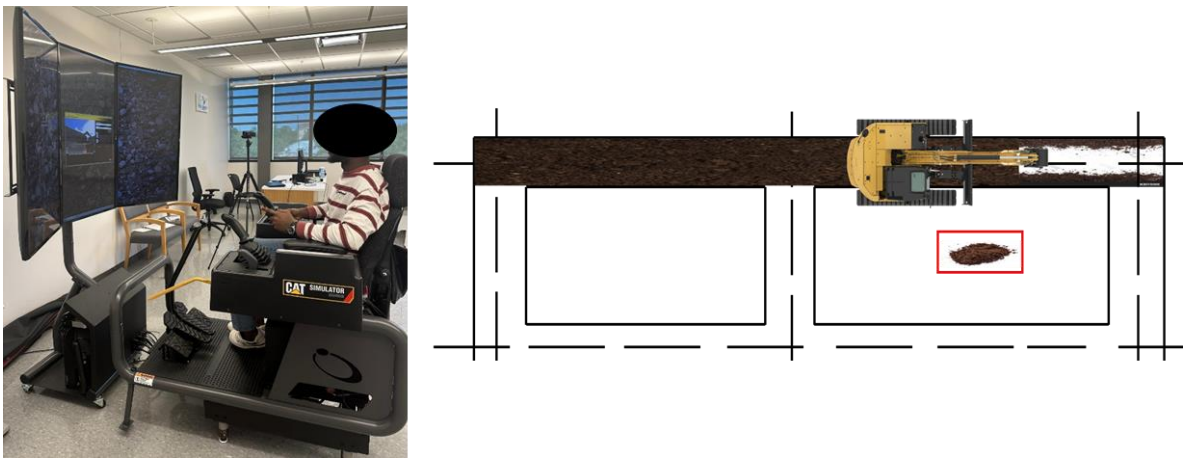


Figure 3-3. CAT Excavator Setup and Digging Exercise

3.4 Video Analysis Pipeline

Rather than extracting individual frames locally and forwarding them to a vision model, the system leverages the native video-understanding capability of the Gemini API. The video recording provided as a URL (e.g., a YouTube link to the simulator session) is submitted directly to the Gemini 2.5 Pro model via the Google GenAI client. A VideoMetadata descriptor accompanies each request, specifying a frame-sampling rate (default: 1 FPS) and optional start and end time offsets that constrain the analyzed segment of the recording. The model processes the video stream internally, eliminating the need for local frame decoding, encoding, or transmission, and enabling temporal reasoning over the full session in a single inference call. This approach was preferred over local frame extraction for two reasons. First, submitting the video as a continuous stream preserves long-range temporal dependencies that would be lost when isolated frames are classified independently; the model can observe how operator behavior evolves over time, for example, whether an initially smooth digging motion progressively degrades into jerky corrections rather

than producing a point-in-time classification for each frame. Second, delegating frame sampling to the API reduces engineering complexity and data-transfer overhead, as no intermediate frame storage, base64 encoding, or batched API calls are required. The model is prompted with a domain-specific system instruction directing it to identify and temporally annotate individual excavation cycles within the recording. The instruction specifies the expected output format: a structured Markdown report containing, among other analytical content, a table of excavation cycle timestamps. Model temperature is configured via a per-prompt TOML file to balance determinism against contextual flexibility.

The structured Markdown report returned by the video-analysis model is parsed to extract excavation cycle data. A regular-expression pattern matches table rows encoding four fields per cycle: the cycle number, start timestamp, end timestamp, and a brief observational note. Start and end times are converted to seconds, and cycle duration is computed as their difference. The resulting list of cycle records, each containing a cycle number, start time, end time, duration, and note, constitutes the primary input to the downstream cycle-metrics stage. This extraction step is entirely deterministic, ensuring that identical model outputs always yield identical structured cycle data.

3.5 Temporal Behavior Analysis

To get interpretable insights into operators' control, we analyzed the temporal behavior of joystick input, which is recorded while performing excavation tasks. Unlike outcome-based performance metrics (e.g., productivity or fuel consumption), this analysis focuses on how well operators can control, capturing simultaneity and coordination directly from joystick time-series data. To achieve these insights, we used three formulations: the Pairwise Simultaneity Index, N-or-More Control Usage, and the Bimanual Coordination Score (BCS). Joystick signals are recorded for the four primary control axes corresponding to excavator operation: left joystick horizontal and vertical axes (L_x, L_y) and right joystick horizontal and vertical axes (R_x, R_y). Each axis represents a distinct machine function (e.g., swing, stick, boom, bucket). Before temporal analysis, each joystick signal is preprocessed and converted into a binary activation sequence $A_i(t)$, where

$$A_i(t) = \begin{cases} 1, & \text{if } |c_i(t)| \geq \theta \\ 0, & \text{otherwise} \end{cases}$$

Here, $c_i(t)$ denotes the continuous joystick signal for axis *i* at time t, and θ is a predefined activation threshold value used to filter out intentional control inputs from noise-level motion. We define the matrices used as follows:

The Pairwise Simultaneity Index (SI) measures how frequently two joystick axes are active at the same time, relative to the total time either axis is active. For any pair of control axes *i* and *j*, we define the simultaneity index as

$$SI(i,j) = \frac{\sum_t A_i(t)A_j(t)}{\sum_t 1\{A_i(t) = 1 \vee A_j(t) = 1\}}$$

The numerator counts the number of time steps at which both controls are simultaneously active, while the denominator counts the number of time steps at which at least one of the two controls is active. This formulation normalizes the simultaneity of two controls by overall usage, allowing meaningful comparison across operators with different activity levels. In the context of excavator operation, higher simultaneity indicates the ability to blend two controls, such as coordinating boom lift with bucket curl, within the same temporal window.

N-or-More Control Proportion formulation quantifies the proportion of time at which at least N controls were active simultaneously. At each time step, the total number of active controls is computed as:

$$N(t) = \sum_{i=1}^{4} A_i(t)$$

The proportion of time during which at least *n* controls are used simultaneously is then defined as:

$$P_{\geq n} = \frac{\sum_t 1\{N(t) \geq n\}}{\sum_t 1\{N(t) \geq 1\}} \times 100$$

This metric captures the operator's tendency to perform compound movements involving multiple degrees of freedom. For excavator tasks, expert operation typically involves higher proportions of triple- or quadruple-control usage, whereas novice operation is often dominated by isolated or sequential control actions.

Finally, the Bimanual Coordination Score (BCS) quantifies the coordination between the left and right hand. We define BCS as the average of the four cross-hand simultaneity indices:

$$BCS = \frac{SI(L_x, R_x) + SI(L_x, R_y) + SI(L_y, R_x) + SI(L_y, R_y)}{4}$$

This formulation explicitly measures the ability of the operator to simultaneously use both hands during task execution, independent of which specific control axes are used. By aggregating cross-hand interactions, the BCS serves as a robust indicator of bimanual coordination and overall operator control skills. Collectively, these three formulations translate raw joystick time-series data into interpretable metrics that can be used to provide feedback to improve the control skills of the operator.

3.6 Simulation Report Extraction

Quantitative performance metrics are extracted from the simulator-generated PDF report using a deterministic, regex-based approach. The trial identifier is first parsed from the input video filename via a regular expression matching alphanumeric identifiers. The corresponding PDF file is then located and its textual content extracted using the pypdf library. Four metrics are recovered through targeted regular expressions: material productivity (m^3/hr), session fuel consumption (L), and superstructure swing time to the left and right (seconds). Both raw-second and HH:MM:SS timestamp formats are supported for the swing-time fields, with automatic normalization to

decimal seconds. This stage is entirely rule-based, ensuring reproducible metric extraction independent of model stochasticity.

3.7 Visual Chart and Cycle Time Analysis

Following telemetry analysis, visual interpretation is performed on the two graphical outputs associated with each trial: the SI heatmap (a color-coded matrix image) and the simultaneous control-usage bar chart. Both images are read from disk, base64-encoded, and submitted as a single multimodal prompt to the Gemini 3 Pro Preview model (temperature = 0.3). Submitting both charts within a single prompt rather than independently allows the model to identify cross-chart patterns; for example, relating a low pairwise SI value in the heatmap to an under-represented control combination in the bar chart, an association that would be difficult to detect from either visualization in isolation. The system prompt instructs the model to identify salient coordination patterns, highlight anomalies relative to the expert benchmarks defined in Section 3.4, and produce a structured Markdown-formatted visual analysis report. The resulting text is subsequently converted to HTML during the report-assembly stage. As with the telemetry analysis stage, a rule-based fallback module produces a default analysis in the event of an API failure.

Excavation cycle durations extracted during the video-analysis phase (Section 3.3.2) are analyzed deterministically, without invoking a language model. Given a sequence of n cycle durations, the mean cycle time μ, variance σ^2, and standard deviation σ are computed using standard descriptive statistics. A consistency score is then derived from the coefficient of variation (CV) and clamped to the interval [0, 100]:

$$d = (d_1, d_2, \ldots, dn)$$

$$C = max(0, 100 \times (1 - \sigma / \mu))$$

A score of 100 indicates perfectly uniform cycle timing, whereas lower values reflect increasing temporal variability. A target cycle time of $T_{target} = 20$ seconds, established from expert benchmark data, is used to compute an efficiency percentage $E = (T_{target} / \mu) \times 100$. To assess within-session performance trends, the sequence of cycle durations is partitioned at the midpoint. Let μ_{first} and μ_{second} denote the means of the first and second halves, respectively. The relative change is defined as $\Delta = (\mu_{second} - \mu_{first}) / \mu_{first} \times 100$. A relative change exceeding +5% is classified as a *declining* trend, a change below −5% as an *improving* trend, and intermediate values as *stable*. A cycle-time trend chart is generated programmatically using Plotly, incorporating reference lines for the target and observed mean cycle times, and is embedded in the final report as a base64-encoded PNG image.

3.8 Insights Generation and Report Assembly

The outputs of the cycle-metrics, joystick-analytics, and chart-analysis stages are synthesized into a holistic operator assessment. A structured prompt is composed, presenting the key quantitative findings from each preceding stage, including cycle-time statistics, the BCS, SI matrix highlights, control-usage distributions, and chart-derived observations together with their respective expert benchmarks. Consolidating all upstream results within a single prompt enables the model to identify cross-domain patterns, for example, correlating high cycle-time variance with low bimanual coordination that would not be apparent from any individual stage output in isolation. This prompt is submitted to the Gemini 3 Pro Preview model (temperature = 0.4) with

constrained JSON output. The model is required to populate a predefined schema comprising six fields:

(1) a *pattern_recognition* object with sub-fields for control, timing, and efficiency patterns;
(2) a *training_recommendations* array of actionable training prescriptions;
(3) an *overall_assessment* narrative;
(4) a proficiency_level classification drawn from {Beginner, Intermediate, Advanced, Expert};
(5) an *estimated_training_hours* value; and
(6) a *next_focus_areas* array listing priority skill-development domains.

Should the model call fail, a deterministic heuristic fallback derives proficiency estimates by combining the consistency score and BCS according to predefined threshold rules, ensuring that a complete report is always produced. The final stage integrates the outputs of all preceding agents into a single, self-contained HTML operator training report. A Jinja2 template engine renders a pre-designed HTML template, injecting operator metadata including name, equipment type, exercise date, and session duration at render time. Quantitative fields are enriched with performance-class labels excellent, good, satisfactory, or needs improvement derived from threshold comparisons against the benchmark values defined in Sections 3.4 and 3.6. Markdown narrative outputs from the chart-analysis and video-analysis agents are converted to equivalent HTML through a lightweight, line-by-line renderer. The decision to produce a self-contained HTML document with all chart images embedded as base64 data URIs and no external asset dependencies was motivated by considerations of portability: reports can be viewed in any web browser, archived, or distributed via email without requiring access to a hosting server or the original data files.

4.0 Results

The temporal behavior analysis framework revealed distinct patterns in joystick control coordination between expert and novice operators, providing quantitative evidence of skill-based differences in excavator operation strategies. The mathematical formulations applied to the joystick signal data captured fundamental variations in control timing, coordination complexity, and bimanual synchronization that characterize different proficiency levels.

4.1 Pairwise Simultaneity Index Analysis

The pairwise simultaneity index (SI) analysis revealed clear and progressive differences in control coordination patterns across novice, intermediate, and expert operators, as illustrated in Table 4.1 and radar chart in Figure 4.1. Expert operators demonstrated substantially higher SI values across all control pair combinations, with the most prominent coordination observed between arm and boom controls (mean SI = 0.542), indicating sophisticated multi-axis control strategies during excavation cycles. The bucket-boom and bucket-arm pairings yielded mean SI values of 0.378 and 0.371 respectively, reflecting the expert group's ability to orchestrate complex hydraulic movements simultaneously. Swing-involved pairs, while lower than the other combinations, showed meaningful coordination at the expert level, with swing-arm (0.276), swing-boom (0.239), and swing-bucket (0.181) reflecting strategic integration of rotational movement into digging operations. Intermediate operators exhibited a distinct coordination profile that differed meaningfully from both groups. While their swing-involved pairs remained comparably low to novice levels (swing-bucket: 0.058, swing-arm: 0.063, swing-boom: 0.043), their working attachment coordination showed notable development, with bucket-arm, arm-boom, and bucket-boom averaging 0.269, 0.257, and 0.254 respectively. This pattern suggests that intermediate skill development is characterized primarily by improved coordination among the three working attachment controls, while swing integration remains underdeveloped. Novice operators displayed the lowest simultaneity indices across all control pairs, consistent with a predominantly sequential control approach. The highest coordination within this group occurred between bucket and arm controls (mean SI = 0.181), representing roughly half the coordination observed at the expert level for the same pair. Arm-boom and bucket-boom pairings averaged 0.125 and 0.144 respectively, indicating limited ability to manage multiple hydraulic functions concurrently. Particularly striking were the near-zero coordination values involving swing operations, with swing-bucket (0.046), swing-arm (0.081), and swing-boom (0.085), suggesting that novice operators largely isolate rotational movement from other excavator functions and rely instead on discrete, single-axis inputs throughout the excavation cycle.

Table 4.1. Pairwise Simultaneity Index (SI)

Control Pair	Novice	Intermediate	Expert
Arm – Boom	0.1251	0.2570	0.5422
Bucket – Boom	0.1442	0.2537	0.3775
Bucket – Arm	0.1806	0.2686	0.3709
Swing – Arm	0.0814	0.0625	0.2764
Swing – Boom	0.0854	0.0431	0.2389
Swing – Bucket	0.0461	0.0578	0.1811

The temporal signal plots in Figure 4.2 illustrate coordination patterns over a 40-second excavation sequence. The expert operator's stick and bucket signals show smooth, sustained overlapping activation, indicated by continuous blue-shaded regions. Signal transitions remain controlled with consistent amplitude ranges, reflecting stable and deliberate control behavior. The persistent overlap throughout the timeline aligns with the high simultaneity indices observed in the heatmap, particularly during the loading and dumping phases.

In contrast, the novice operator's signals exhibit fragmented and irregular activation patterns, with brief purple-shaded intervals indicating limited simultaneous control. The signals show frequent oscillations and variable amplitudes, suggesting corrective and less stable inputs. The intermittent overlap reflects a reliance on sequential, single-axis movements, consistent with the lower simultaneity indices observed in the novice heatmap and indicating reduced coordination efficiency.

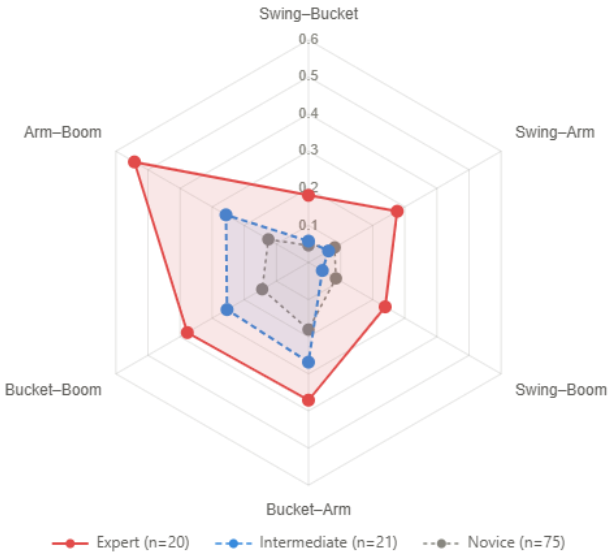


Figure 4-1. Mean Simultaneity Index (SI) by Control Pair and Operator Skill Level

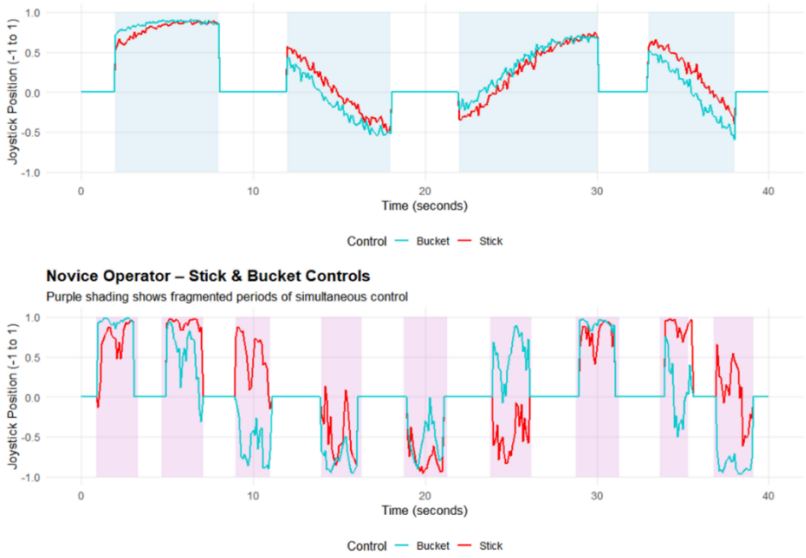


Figure 4-2. Simultaneity Index Visualization

4.2 Bimanual Coordination Score Distribution

The Bimanual Coordination Score (BCS) demonstrated a clear and statistically significant progressive increase across skill levels, as illustrated in Figure 4.3 and summarized in Table 4.2. Expert operators achieved a mean BCS of 0.333 (SD = 0.025), compared to 0.157 (SD = 0.017) for intermediate operators and 0.109 (SD = 0.061) for novice operators. A Kruskal-Wallis test confirmed that group differences were statistically significant ($H = 62.43$, $p < 0.001$), with pairwise Mann-Whitney tests indicating significant separation between all three groups ($p < 0.001$ in all comparisons). Beyond the differences in central tendency, the distributional shape of BCS scores revealed an important pattern across skill levels. Expert operators formed a tight, high-valued cluster, reflecting consistent bimanual coordination across individuals in that group. Intermediate operators similarly showed low variability, occupying a narrow band in the mid-range. Novice operators, by contrast, exhibited the widest spread, spanning from near-zero to values approaching intermediate levels, suggesting that bimanual coordination among novices is highly variable and not yet stabilized. Notably, a small number of novice operators produced BCS values in the range of 0.28 to 0.32, sitting well above the bulk of their group and approaching the lower boundary of the expert distribution. These individuals may represent novices with prior exposure to related equipment or stronger innate spatial coordination, and warrant attention in future proficiency development research. This divergence in distributional shape, rather than mean differences alone, underscores the developmental nature of simultaneous control and its sensitivity as a skill discriminator.

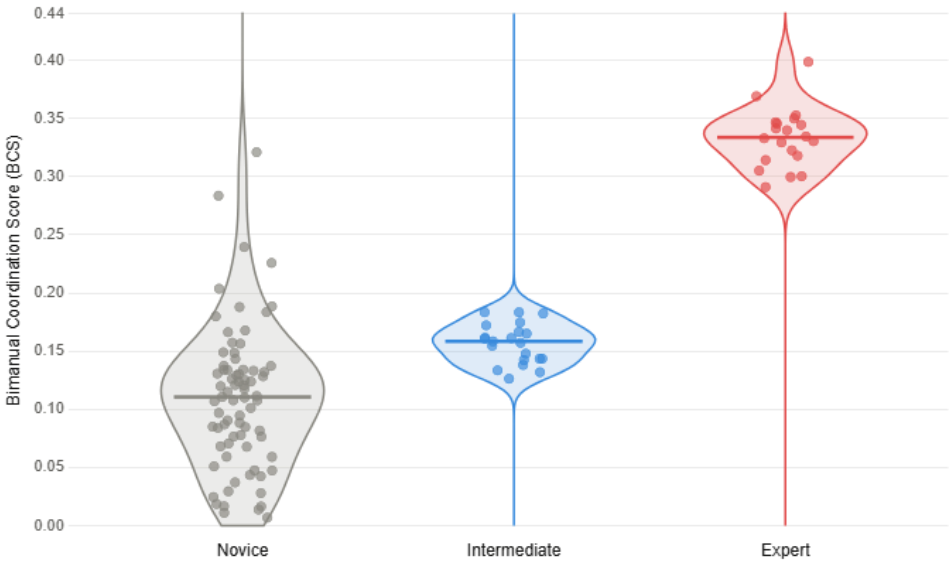


Figure 4-3. Bimanual Coordination Score. Horizontal Line Denotes Group Median

These findings have significant implications for operator training and assessment. The quantitative thresholds established by the BCS analysis enable objective classification of operator skill levels, with scores below 0.14 indicating novice performance, scores between 0.14 and 0.25 suggesting intermediate proficiency, and scores above 0.25 characteristic of expert operation. Furthermore, the substantial gap between intermediate and expert BCS values highlights the extended practice and experience required to achieve true operational mastery, suggesting that training programs should emphasize sustained practice of coordinated control movements rather than focusing solely on individual joystick proficiency.

Table 4.2. Mean Bimanual Coordination Score (BCS)

Novice	Intermediate	Expert
0.1093 ± 0.061	0.1566 ± 0.017	0.3333 ± 0.025

4.3 Multi-Control Usage Patterns

The analysis of multi-control usage patterns across skill levels revealed fundamental differences in control complexity management, as illustrated in Figure 4.4. The N or More Control Usage metric quantified the percentage of operational time spent utilizing different numbers of simultaneous controls, providing insights into how operators of varying expertise manage the cognitive and motor demands of multi-axis coordination. Novice operators exhibited a predominantly conservative control strategy, spending only 33.8% of their operational time using two controls simultaneously. This limited dual-control usage primarily occurred during basic excavation movements where simple combinations, such as boom lowering with bucket curl, were absolutely necessary. The three-control configuration accounted for merely 6.3% of novice operational time, indicating significant difficulty in managing triple-axis coordination. Most notably, four-control simultaneous operation was virtually absent at 0.5%, demonstrating that novices essentially avoid the complex coordination required for smooth, efficient excavation cycles. This distribution pattern reflects a cognitive bottleneck where novices must consciously process each control input, limiting their ability to execute multiple movements in parallel. Intermediate operators showed marked improvement in multi-control coordination, with two-control usage increasing to 49.7% of operational time. This near 50% increase over novice performance indicates growing comfort with basic simultaneous movements and the development of fundamental coordination patterns. The three-control configuration usage remained relatively modest at 6.8%, showing only marginal improvement over novice levels. Four-control usage remained negligible at 0.1%, suggesting that even intermediate operators struggle with the highest levels of coordination complexity. The pattern indicates that intermediate skill development primarily involves mastering two-control combinations rather than advancing to more complex multi-axis coordination. Expert operators demonstrated a dramatic shift in control strategy, spending 80.3% of their operational time utilizing two controls simultaneously, more than double the novice rate and significantly exceeding intermediate performance. This extensive dual-control usage reflects the automation of basic coordination patterns, allowing experts to maintain continuous multi-axis movement throughout most of the excavation cycle. Remarkably, experts utilized three-control configurations for 35.7% of their operational time, representing a five-fold increase over novice and intermediate operators. This substantial three-control usage enables experts to execute complex maneuvers such as simultaneous boom lowering, stick extension, and bucket curl during loading operations. Even four-control simultaneous operation reached 4.9% in expert performance, nearly ten times the novice rate, indicating mastery of the most demanding coordination challenges in excavator operation. The progression from novice to expert reveals a fundamental transformation in control strategy. While novices operate in a predominantly sequential mode with occasional dual-control coordination, experts maintain near-continuous multi-axis control with frequent periods of high-complexity coordination. The increase in three and four-control usage at the expert level suggests a qualitative shift in motor control organization, where complex coordination patterns become integrated motor programs rather than consciously managed individual movements.

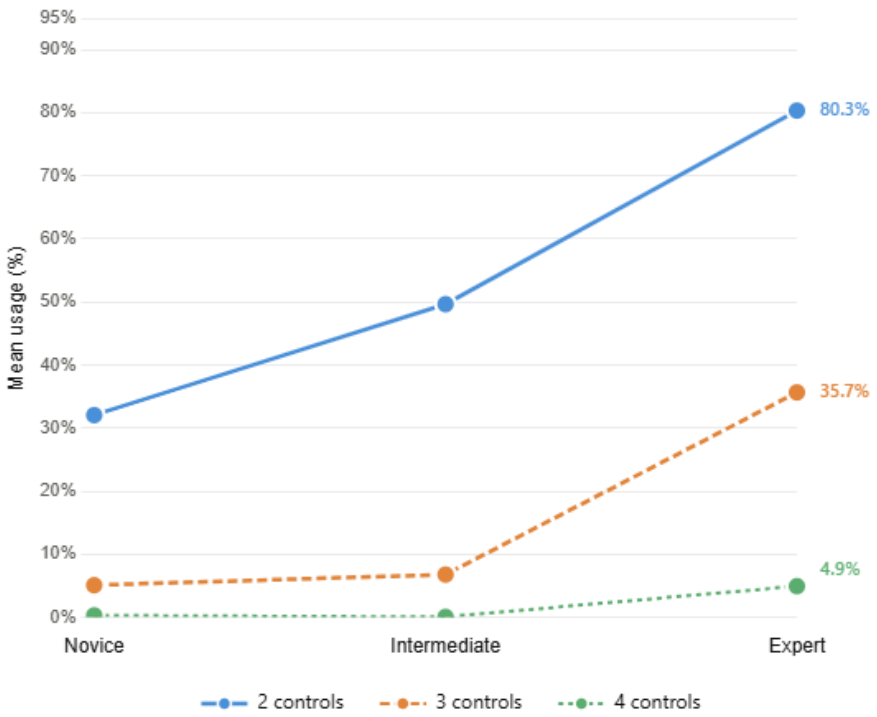


Figure 4-4. Control Usage (%) by Operator Skill Level

The distribution patterns also reveal strategic differences in operational approach. Novices' low multi-control usage results in discrete, segmented movements that extend cycle times and reduce operational fluidity. Intermediate operators' increased two-control usage improves basic operational flow but lacks the sophisticated three-control coordination necessary for optimal performance. Experts' high utilization across all control configurations enables smooth, continuous movements that minimize cycle time while maximizing precision and efficiency. These findings have important implications for training progression and assessment. The data suggests that advancing from novice to intermediate primarily requires mastering two-control coordination, while the transition from intermediate to expert demands developing proficiency in three and four-control configurations. Training programs should therefore implement staged progression, first establishing robust two-control patterns before introducing more complex multi-axis coordination exercises. Furthermore, the dramatic differences in multi-control usage patterns provide objective benchmarks for skill assessment, with two-control usage above 70% and three-control usage above 25% serving as indicators of expert-level proficiency.

4.4 **AI-Generated Operator Training Reports**

This section presents paired case studies demonstrating how the proposed framework translates control-signal analytics into structured operator training reports. The generated reports represent the final operational layer of the framework, converting joystick telemetry, temporal coordination metrics, and performance indicators into interpretable training feedback. Rather than functioning as descriptive summaries, the reports operationalize the mathematical constructs developed in this study to support diagnostic reasoning and targeted skill development.

4.5 Case Study A: Advanced-Level Operator

The Advanced-level operator (Figure 4-5) achieved an average cycle time of 16.5 seconds against a 20.0-second benchmark, a consistency score of 79.2%, and a BCS of 0.333, reflecting strong foundational proficiency. The SI heatmap reveals a concentrated coordination signature among the digging attachment controls, with the Arm–Boom pair yielding the highest value (SI = 0.582), followed by Bucket–Boom (0.425) and Bucket–Arm (0.386). These values indicate well-developed core digging mechanics. However, all Swing-involved pairs were markedly lower , Swing–Bucket (0.147), Swing–Boom (0.217), pointing to a specific coordination ceiling rather than a generalized deficiency. This pattern is corroborated by the control usage distribution: while dual-control usage was exceptional at 84.3%, well above the 65% benchmark, and triple-control usage met the 35% threshold at 36.5%, full four-control engagement reached only 4.8% against a 10% benchmark. The framework interprets this signature as a "Stop-and-Go Swing" behavior, the operator effectively pauses cab rotation rather than blending it into the digging cycle. Accordingly, the report prescribes targeted drills: "Dump-on-the-Fly" exercises to encourage bucket actuation during the final degrees of swing rotation, and "The Waiter" drill to develop simultaneous boom lift and swing, directly addressing the low Swing–Boom SI value.

4.6 Case Study B: Novice-Level Operator

The Beginner-level operator (Figure 4-6) presents a fundamentally different coordination profile. The average cycle time of 35.12 seconds, nearly double the benchmark, combined with a consistency score of 59.1% and a declining cycle trend across the session indicate both inefficiency and cognitive fatigue. The BCS of 0.149 reflects developing but unstabilized bimanual coordination. The SI heatmap is uniformly pale, with Bucket–Arm as the relative strength at 0.283, while all other pairs remain below 0.150, and Swing–Bucket registers near zero at 0.042. The control usage distribution confirms a "Sequential Isolation" pattern: dual-control usage stands at 38.4%, well below the 65% benchmark, triple-control at 3.4%, and full four-control usage at 0.0%. Rather than prescribing advanced integration drills, the framework shifts its reasoning accordingly, recommending "Flowing the Needle" exercises to build smooth simultaneous Bucket–Arm blending, diagonal lift practice to break the lift-then-swing habit, and slow-motion multi-function cycles to develop the neural pathways for triple-control coordination. The estimated training hours required reflect this gap, at 45 hours compared to 15 hours for the Advanced operator.

4.7 Comparative Diagnostic Utility: Behavioral Patterns Across Skill Levels

Direct comparison of these case studies reveals systematic variations in coordination behavior that align with established models of psychomotor skill acquisition. The framework utilizes the same mathematical constructs,specifically the Bimanual Coordination Score (BCS) and Simultaneity Index (SI), as an interpretive substrate to differentiate between skill profiles. While the advanced operator demonstrated sustained multi-axis control, the novice profile was characterized by a reliance on sequential movements and a total absence of four-control engagement. For instance, the advanced operator’s BCS (0.333) significantly exceeded that of the novice (0.149), providing a quantitative metric for cross-hand synchronization. These data points suggest that temporal coordination metrics capture nuanced behavioral distinctions that traditional outcome-based indicators, such as total cycle time, often overlook. Consequently, the framework mirrors the adaptive reasoning of an expert trainer by mapping these identified bottlenecks to specific, evidence-based corrective interventions.

Operator Name:	Operator (Video: Ddm6NSjKCyk)
Equipment:	Excavator
Exercise Date:	2025-12-10
Session Duration:	N/A
Analysis System:	Gemini 2.5 Pro Multi-Agent System

□ SIMULATION PERFORMANCE METRICS

Report ID: B6

PRODUCTIVITY	FUEL BURNED	SWINGING LEFT	SWINGING RIGHT
661.98 m³/hr	1.41 Liters	44.0 seconds	43.0 seconds

1. EXECUTIVE SUMMARY

This report provides comprehensive analysis of operator performance including productivity metrics, advanced control coordination analytics, and safety compliance. Key findings and recommendations are highlighted throughout.

Overall Assessment

The operator demonstrates strong Advanced-level proficiency, characterized by aggressive cycle times significantly beating the target (16.5s vs 20.0s) and excellent bimanual coordination (0.333). Their mastery of the primary digging geometry (Boom and Arm coordination) is a standout strength, allowing for rapid material acquisition. The high Dual Control usage confirms they have moved well beyond novice single-function operation. However, the analysis reveals a specific technical ceiling regarding the Swing function. The operator acts as a 'Stop-and-Go Swinger,' isolating cab rotation from the rest of the cycle. While they are currently fast, this lack of fluidity prevents them from reaching Expert status and places unnecessary structural stress on the machine due to jerky transitions. Focusing on blending the swing with lift and dump functions will unlock the next tier of performance.

2. PERFORMANCE METRICS

2.1 Cycle Time Analysis

Metric	Value	Benchmark	Performance
Average Cycle Time	16.5s	20.0s	Above Target
Cycle Time Variance	11.83	<15%	Good
Consistency Score	79.2%	>70%	Good
Cycle Time Trend	Improving (faster over time)		

3. ADVANCED CONTROL ANALYTICS

Multi-Joystick Coordination

The Simultaneity Index (SI) Matrix shows how frequently different controls are used together. Higher values indicate better coordination between specific control pairs.

Control Pair	Function Mapping	Simultaneity Index (SI)
Left X + Right X	(Swing + Bucket)	0.147
Left X + Right Y	(Swing + Boom)	0.217
Left Y + Right X	(Arm + Bucket)	0.386
Left Y + Right Y	(Arm + Boom)	0.582

Bimanual Coordination Score (BCS)

Overall BCS: 0.333
Interpretation: Moderate coordination; operator is beginning to blend movements.

Simultaneous Control Usage Distribution

Control Configuration	% of Active Time	Expert Benchmark	Status
Single Control (≥1)	100%	100%	N/A
Dual Control (≥2)	84.3%	>65%	Strong
Triple Control (≥3)	36.5%	>35%	Average
Full Control (4)	4.8%	>10%	Needs Improvement

Key Finding: While the operator has mastered basic dual-function movements, the low utilization of full control (4 functions) suggests they struggle with complex, compound maneuvers and should focus on integrating all axes simultaneously.

4. VISUAL CHART ANALYSIS

Simultaneity Index (SI) Heatmap



Control Usage Distribution



Based on the visual data provided in the Simultaneity Index Heatmap and the Simultaneous Control Usage chart, here is the comprehensive analysis report.

Visual Chart Analysis

Simultaneity Index Heatmap Insights

Overall Coordination Patterns

The heatmap reveals a distinct "clustering" of coordination among the digging attachment controls (Boom, Arm, and Bucket), represented by the darker blue shades in the bottom-right quadrant of the matrix. In contrast, the "Swing" function (top row and left column) is consistently lighter in color across all pairings. This indicates the operator is highly proficient at manipulating the arm geometry but tends to isolate the swing function rather than blending it seamlessly with other movements.

Specific Control Pair Analysis

- **Boom + Arm**: SI = **0.582** - This is the darkest blue square on the chart, indicating the strongest coordination. This represents the core digging motion (hoisting the boom while curling the arm in). This high value suggests a smooth, efficient power stroke during excavation.
- **Boom + Bucket**: SI = **0.425** - The medium-dark blue indicates good competency in maintaining bucket angle while raising or lowering the boom. This is critical for preventing spillage while lifting a load.
- **Bucket + Swing**: SI = **0.147** - This is the lowest value (lightest color) on the entire map. It indicates the operator rarely dumps or curls the bucket while the cab is rotating. They likely complete the swing, stop, and *then* dump the bucket, which introduces inefficiency.

Coordination Strengths

1. **Boom & Arm (0.582):** Exceptional coordination in the primary digging cycle.
2. **Boom & Bucket (0.425):** Strong ability to keep the load level during elevation changes.

Coordination Weaknesses

1. **Bucket & Swing (0.147):** The operator is treating these as sequential rather than simultaneous actions.
2. **Boom & Swing (0.217):** The relatively light color suggests the operator is not consistently lifting the load *while* swinging toward the dump pile, often waiting to lift before rotating.

Control Usage Distribution Analysis

Current Performance Levels

Based on the bar chart, the operator's usage percentages are:

- **Dual Control (≥2): 84.3%** - This is significantly higher than the 65% benchmark, indicating the operator almost never moves a single function in isolation.
- **Triple Control (≥3): 36.5%** - This aligns perfectly with (and slightly exceeds) the 35% benchmark, showing solid intermediate proficiency.
- **Full Control (4): 4.8%** - This is roughly half of the 10% benchmark, indicating a hesitation to engage all functions at once.

Visual Trends

The chart shows a healthy foundation but a steep drop-off at the highest complexity level. The first bar (84.3%) is massive, dominating the visual field, which proves the operator has moved past "single-function" thinking. However, the drop from Triple (36.5%) to Full (4.8%) is visually dramatic, suggesting a specific skill ceiling has been hit where the operator cannot mentally process the fourth dimension of movement (usually the Swing) alongside the other three.

Skill Level Assessment

Advanced.

While the "Full Control" numbers are low, the "Dual Control" numbers are exceptional (nearly 20% above benchmark), and "Triple Control" is on target. The operator has mastered the digging mechanics but has not yet achieved the "fluidity" of an Expert who blends all four movements seamlessly.

Key Visual Observations

Pattern Recognition

The combination of the two charts tells a clear story: **The "Stop-and-Go" Swinger.**

The Heatmap shows the Swing axis is the "coldest" part of the operator's technique. Correlating this with the Bar Chart, the missing 5% in the "4 controls" category is almost certainly the lack of adding "Swing" to the "Boom/Arm/Bucket" trio. The operator digs efficiently (high Boom/Arm SI), but likely pauses slightly before initiating the swing, or stops the digging motion to focus on the swing.

Comparison to Expert Patterns

An expert heatmap would show a more uniform distribution of medium-to-dark blues. Specifically, the **Swing + Boom** and **Swing + Arm** squares would be much darker (closer to 0.4 - 0.5) in an expert profile, representing the "swing-and-lift" technique. Additionally, an expert's bar chart would show a "4 controls" bar closer to the 10-12% range.

Technical Recommendations

1. **"The Waiter" Drill (Swing + Boom):** Practice lifting a bucket of water (or loose material) while swinging 90 degrees, aiming to reach full dump height exactly when the swing terminates. This targets the low 0.217 SI score.
2. **"Dump-on-the-Fly" (Swing + Bucket):** Practice initiating the bucket dump while the machine is still in the last 15 degrees of the swing rotation, rather than waiting for a full stop. This addresses the 0.147 SI weakness.

Summary

This operator demonstrates excellent fundamental digging mechanics, evidenced by a very high 0.582 correlation between Boom and Arm and exceptional Dual Control usage (84.3%). However, visual analysis identifies a specific bottleneck in **Swing integration**; the operator tends to isolate the rotation of the cab from the digging attachment movements. Training should focus exclusively on blending the swing function with boom lift and bucket dumping to push this operator from Advanced to Expert status.

5. AI-GENERATED INSIGHTS & RECOMMENDATIONS

5.1 Pattern Recognition Analysis

- **Control Pattern:** The operator displays exceptional coordination in the digging phase (Boom/Arm SI 0.582) and high Dual Control usage (84.3%), indicating strong fundamental mechanics. However, there is a distinct isolation of the Swing function, evidenced by the low Swing+Bucket SI (0.147) and low Full Control usage (4.8%). The operator effectively treats digging and swinging as sequential rather than simultaneous events.
- **Timing Pattern:** Cycle times are aggressive (16.5s average vs 20.0s target) and trending faster, showing high operator confidence. However, the Consistency Score of 79.2% suggests that the 'Stop-and-Go' nature of their swing technique creates variable micro-pauses. The operator is rushing the dig phase to compensate for the lack of fluidity in the swing phase.
- **Efficiency Pattern:** The improving trend indicates a quick learning curve or successful warm-up period. However, the sequential swing technique creates kinetic inefficiency; stopping the swing completely before dumping requires more energy to restart motion compared to a fluid, continuous momentum dump.

5.2 Personalized Training Recommendations

- Implement 'The Waiter' Drill: Focus on lifting the boom while simultaneously swinging 90 degrees, aiming to reach maximum height exactly as the swing terminates to address the low Swing+Boom correlation.
- Execute 'Dump-on-the-Fly' Exercises: Instruct the operator to initiate the bucket curl/dump while the machine is still in the final 15 degrees of rotation to eliminate the static pause identified in the heatmap.
- Practice 'Blind Swing' Drills: Have the operator perform cycles focusing purely on the sensation of the swing brake to encourage smoother deceleration and blending, aiming to raise Full Control usage toward the 10% target.

6. SUMMARY & NEXT STEPS

6.1 Overall Assessment

The operator demonstrates strong Advanced-level proficiency, characterized by aggressive cycle times significantly beating the target (16.5s vs 20.0s) and excellent bimanual coordination (0.333). Their mastery of the primary digging geometry (Boom and Arm coordination) is a standout strength, allowing for rapid material acquisition. The high Dual Control usage confirms they have moved well beyond novice single-function operation. However, the analysis reveals a specific technical ceiling regarding the Swing function. The operator acts as a 'Stop-and-Go Swinger,' isolating cab rotation from the rest of the cycle. While they are currently fast, this lack of fluidity prevents them from reaching Expert status and places unnecessary structural stress on the machine due to jerky transitions. Focusing on blending the swing with lift and dump functions will unlock the next tier of performance.

6.2 Certification Readiness

Current Proficiency Level:	Advanced
Estimated Training Hours Required:	15.0 hours

6.3 Next Training Session Focus Areas

- Blending Swing rotation with Boom Lift
- Initiating Bucket Dump during Swing deceleration
- Increasing Full Control (4-function) simultaneity

Figure 4-5. *Sample AI-generated operator training report :Advanced-level case study*

OPERATOR TRAINING REPORT

Post-Exercise Performance Analysis with Advanced Control Analytics

Operator Name:	Operator (Video: dDpmny4Ui5s)
Equipment:	Excavator
Exercise Date:	2025-12-10
Session Duration:	N/A
Analysis System:	Gemini 2.5 Pro Multi-Agent System

□ SIMULATION PERFORMANCE METRICS

Report ID: 52

PRODUCTIVITY	FUEL BURNED	SWINGING LEFT	SWINGING RIGHT
531.94	1.86	32.0	44.0
m³/hr	Liters	seconds	seconds

1. EXECUTIVE SUMMARY

This report provides comprehensive analysis of operator performance including productivity metrics, advanced control coordination analytics, and safety compliance. Key findings and recommendations are highlighted throughout.

Overall Assessment

The operator is currently functioning at a Beginner level, characterized by a 'step-by-step' operating style. The data reveals a 'Sequential Isolation' behavior where the operator manipulates the machine's joints individually rather than as a fluid extension of their body. While there is a foundational understanding of the primary digging mechanic (Bucket and Arm coordination is their strongest point), the near-total absence of Triple (3.4%) and Full (0%) control usage indicates a lack of ingrained muscle memory. The operator is currently thinking through every movement, leading to mental fatigue, which explains the declining trend in speed over the 8 cycles. To progress, the operator must move away from accuracy-obsessed, robotic movements and embrace 'blending.' The current average cycle time of 35.12s is nearly double the target, largely because momentum is lost every time a control input stops. The focus must shift from precision to fluidity; once the operator learns to overlap the Swing phase with Boom and Arm movements, cycle times will naturally decrease, and the cognitive load will drop, allowing for better consistency.

2. PERFORMANCE METRICS

2.1 Cycle Time Analysis

Metric	Value	Benchmark	Performance
Average Cycle Time	35.12s	20.0s	Below Target
Cycle Time Variance	206.41	<15%	High Variance
Consistency Score	59.1%	>70%	Satisfactory
Cycle Time Trend	Declining (slower over time)		

3. ADVANCED CONTROL ANALYTICS

Multi-Joystick Coordination

The Simultaneity Index (SI) Matrix shows how frequently different controls are used together. Higher values indicate better coordination between specific control pairs.

Control Pair	Function Mapping	Simultaneity Index (SI)
Left X + Right X	(Swing + Bucket)	0.042
Left X + Right Y	(Swing + Boom)	0.144
Left Y + Right X	(Arm + Bucket)	0.283
Left Y + Right Y	(Arm + Boom)	0.126

Bimanual Coordination Score (BCS)

Overall BCS: 0.149
Interpretation: Developing coordination; operator frequently isolates hand movements.

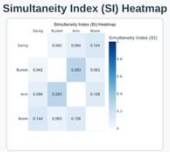
Simultaneous Control Usage Distribution

Control Configuration	% of Active Time	Expert Benchmark	Status
Single Control (≥1)	100%	100%	N/A
Dual Control (≥2)	38.4%	>65%	Needs Improvement
Triple Control (≥3)	3.4%	>35%	Needs Improvement
Full Control (4)	0.0%	>10%	Needs Improvement

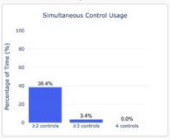
Key Finding: The operator relies heavily on sequential, single-function movements and needs to focus on basic fluidity to begin blending at least two controls simultaneously.

4. VISUAL CHART ANALYSIS

Simultaneity Index (SI) Heatmap



Control Usage Distribution



Here is the comprehensive visual analysis report based on the provided charts.

Visual Chart Analysis

Simultaneity Index Heatmap Insights

Overall Coordination Patterns

The heatmap is predominantly characterized by very light colors (white to pale blue), indicating a generally low level of simultaneous control usage across the board. There are no dark blue regions (values approaching 0.8 or 1.0), which suggests this operator tends to use controls sequentially rather than blending them fluidly. The highest coordination values are clustered around the **Arm** and **Bucket** intersection, while combinations involving **Swing** and **Bucket** show the lowest coordination (near white).

Specific Control Pair Analysis

- **Bucket + Arm**: SI = **0.283** - This is the highest value on the chart, represented by the darkest shade of light blue. This combination represents the primary digging cycle (curling the bucket while pulling the arm in). While it is the operator's strongest pair, a value of 0.283 is still relatively low, suggesting the operator often pauses the arm movement to adjust the bucket, rather than curling smoothly through the pull.
- **Swing + Boom**: SI = **0.144** - This is the second-highest value. This movement enables lifting the load while rotating toward the dump truck or spoil pile. The low value indicates the operator likely lifts the boom first, stops, and then swings, rather than performing a diagonal lift-and-turn motion.
- **Arm + Boom**: SI = **0.126** - This combination is critical for maintaining a flat grade (leveling) or reaching. The low score suggests the operator struggles to adjust boom height dynamically as the arm extends or retracts.

Coordination Strengths

Based on the relative color intensity in the heatmap, the operator's "strengths" (relative to their own baseline) are:

1. **Bucket & Arm** (0.283) - The fundamental digging mechanic.
2. **Swing & Boom** (0.144) - The basic transport mechanic.

Coordination Weaknesses

The following pairs show extremely low SI values (near white), indicating almost zero simultaneous usage:

1. **Swing & Bucket** (0.042) - The operator is unable to adjust the bucket angle while rotating the cab.
2. **Bucket & Boom** (0.063) - The operator rarely curls/dumps the bucket while raising/lowering the boom.

Control Usage Distribution Analysis

Current Performance Levels

Based on the bar chart data:

- **Dual Control (≥2): 38.4%** - This is significantly below the 65% benchmark. The operator spends the majority of their time using only single controls (one axis at a time).
- **Triple Control (≥3): 3.4%** - This is drastically lower than the 35% benchmark. The operator almost never manages three functions at once (e.g., Swing + Boom Up + Arm Out).
- **Full Control (4): 0.0%** - This is below the 10% benchmark. The operator never utilizes all four inputs simultaneously.

Visual Trends

The bar chart exhibits a **steep "cliff" drop-off pattern**. While there is a moderate bar for dual control (38.4%), the usage creates a near-vertical drop to the triple control category (3.4%) and flatlines at zero for full control. This visual trend indicates a cognitive bottleneck; the operator can manage two hands to a limited extent but becomes overwhelmed when a third variable is introduced.

Skill Level Assessment

Beginner

The visual data strongly points to a novice operator. The combination of low dual control usage (under 40%) and the near-total absence of triple control usage indicates an operator who is still "thinking" about each individual move rather than feeling the machine flow. They are operating sequentially (step-by-step) rather than simultaneously.

Key Visual Observations

Pattern Recognition

The unique pattern visible here is the **"Sequential Isolation"** pattern.

- In the heatmap, the isolation is visible because the cells are distinct shades rather than a blended gradient of high activity.
- In the bar chart, the absence of a "tail" (the 3 and 4 control bars) confirms that the operator treats the excavator like a crane—move one joint, stop, move the next—rather than a fluid extension of the body.

Comparison to Expert Patterns

- **Heatmap:** An expert heatmap would feature dark blue squares (0.6+) for **Bucket/Arm** and **Swing/Boom**, and mid-blue tones for **Arm/Boom**. This chart is far too pale.
- **Bar Chart:** An expert chart would show a "stair-step" decline rather than a cliff. We would expect the first bar to be near 70-80%, the second near 40-50%, and the third visible around 10-15%. This operator lacks the "blending" capability of an expert.

Technical Recommendations

1. **"Flowing the Needle" Drill:** To improve the **Bucket + Arm (0.283)** score, have the operator practice long, slow strokes where they must keep the bucket teeth moving at a constant speed while the arm pulls in, forcing them to modulate both joysticks simultaneously without stopping.
2. **Diagonal Lift Practice:** To address the low **Swing + Boom (0.144)** and **Triple Control (3.4%)** scores, practice lifting a load from ground level to a high target at a 45-degree angle. Force the operator to engage the swing *before* the boom reaches full height, prohibiting sequential "lift then turn" movements.

Summary

The visual analysis reveals a **Beginner** operator who relies heavily on sequential inputs rather than fluid, simultaneous movements. The **Simultaneity Index Heatmap** is uniformly low, with the best coordination (Bucket/Arm) only reaching 0.283, while the **Control Usage Chart** shows a critical drop-off in multi-function capability (0% usage of 4 controls). Training should prioritize "blending" exercises to encourage the operator to overlap functions, specifically focusing on combining Swing operations with Boom/Arm movements to increase cycle efficiency.

5. AI-GENERATED INSIGHTS & RECOMMENDATIONS

5.1 Pattern Recognition Analysis

- **Control Pattern:** The operator exhibits a distinct 'Sequential Isolation' pattern, treating the excavator like a crane by moving one joint at a time rather than blending functions. This is evidenced by the 'cliff' drop-off in control usage: 38.4% Dual usage plummets to 3.4% Triple usage and 0% Full usage. The operator struggles significantly with the left hand/right hand coordination required for Swing + Bucket (SI 0.042).
- **Timing Pattern:** There is a concerning declining trend in cycle times, indicating cognitive fatigue or frustration as the session progressed. The high variance between the best cycle (22s) and the worst (69s), combined with a low consistency score of 59.1%, suggests the operator pauses frequently to 'think' about the next move rather than relying on muscle memory.
- **Efficiency Pattern:** The operator faces a cognitive bottleneck when a third variable is introduced. While they have basic grasp of the primary digging cycle (Bucket + Arm), the efficiency collapses during the swing/transport phase because they lift, stop, and then swing, rather than performing a diagonal lift. This sequential approach is the primary driver of the inflated 35.12s average cycle time.

5.2 Personalized Training Recommendations

- Implement 'Flowing the Needle' drills: Practice long, slow strokes focusing solely on maintaining constant velocity while blending Bucket and Arm inputs to improve the 0.283 SI score.
- Diagonal Lift Exercises: Force the operator to engage the Swing function before the Boom reaches 50% height during lifting operations to break the 'lift-then-turn' habit.
- Slow-Motion Multi-Function: Have the operator perform cycles at 25% speed, but with the strict rule that 3 functions must be engaged at all times (even if moving slightly), to build neural pathways for Triple Control.
- Blind-Spot Digging: To reduce visual reliance and cognitive load, have the operator perform simple digging cycles without looking at the controls, encouraging 'feeling' the machine response.

6. SUMMARY & NEXT STEPS

6.1 Overall Assessment

The operator is currently functioning at a Beginner level, characterized by a 'step-by-step' operating style. The data reveals a 'Sequential Isolation' behavior where the operator manipulates the machine's joints individually rather than as a fluid extension of their body. While there is a foundational understanding of the primary digging mechanic (Bucket and Arm coordination is their strongest point), the near-total absence of Triple (3.4%) and Full (0%) control usage indicates a lack of ingrained muscle memory. The operator is currently thinking through every movement, leading to mental fatigue, which explains the declining trend in speed over the 8 cycles. To progress, the operator must move away from accuracy-obsessed, robotic movements and embrace 'blending.' The current average cycle time of 35.12s is nearly double the target, largely because momentum is lost every time a control input stops. The focus must shift from precision to fluidity; once the operator learns to overlap the Swing phase with Boom and Arm movements, cycle times will naturally decrease, and the cognitive load will drop, allowing for better consistency.

6.2 Certification Readiness

Current Proficiency Level:	Beginner
Estimated Training Hours Required:	45 hours

6.3 Next Training Session Focus Areas

- Blending Bucket and Arm inputs for a smooth curl
- Overlapping Swing and Boom for diagonal lifting
- Improving cycle consistency to reduce cognitive fatigue

Report Generated: 2025-12-10 17:35:39
Analysis System: Gemini 2.5 Pro Multi-Agent Training Analysis System v1.0
Data Sources: Video analysis, joystick telemetry, performance metrics database

Figure 4-6. *Sample AI-generated operator training report:Beginner-level case study.*

4.8 Video-Derived Error Detection and Alignment with Expert Review

In addition to joystick and telemetry analysis, the framework incorporates an automated video-snipping module designed to detect and timestamp operational irregularities during excavation tasks. This module monitors deviations in bucket trajectory, abrupt motion transitions, and mismatches between expected and observed task phases. Detected events are extracted as short, time-stamped video segments that serve as localized evidence of performance deficiencies and support both automated evaluation and subsequent expert review. Across the analyzed trials, the system consistently identified operational errors associated with coordination breakdowns and control instability. For example, in one detected event (Figure 4.6), premature backtracking occurred while the bucket remained inside the trench, causing the bucket to contact the trench wall and displace surrounding material. This behavior reflects a mismatch between bucket position and intended retreat motion and is typically associated with insufficient spatial awareness or incomplete coordination between swing and travel controls. A second detected event involved incorrect dump placement during the unloading phase (Figure 4.7), where material was discharged outside the designated dump zone. This misalignment indicates either premature bucket rotation, excessive swing angle, or delayed stabilization prior to dumping. Such errors introduce inefficiencies by requiring corrective repositioning and additional handling effort. Finally, the system identified an episode of jerky and unstable dumping motion (Figure 4.8), characterized by abrupt acceleration and inconsistent hydraulic modulation. This behavior is indicative of overcorrection or poor control finesse during the unloading phase and is commonly observed in operators with limited coordination stability. These automatically detected issues were subsequently reviewed by experienced trainers, who independently identified the same operational deficiencies while analyzing the full-length recordings. The agreement between automated detection and expert interpretation demonstrates that the video-snipping module captures behaviorally meaningful errors and provides reliable visual evidence to support instructor-level evaluation. By linking temporal coordination metrics with visually observable performance events, the video-based detection component strengthens the interpretability of the framework and enables objective documentation of training deficiencies. This capability extends the framework beyond numerical performance assessment by providing direct, verifiable evidence of operational behavior during task execution.

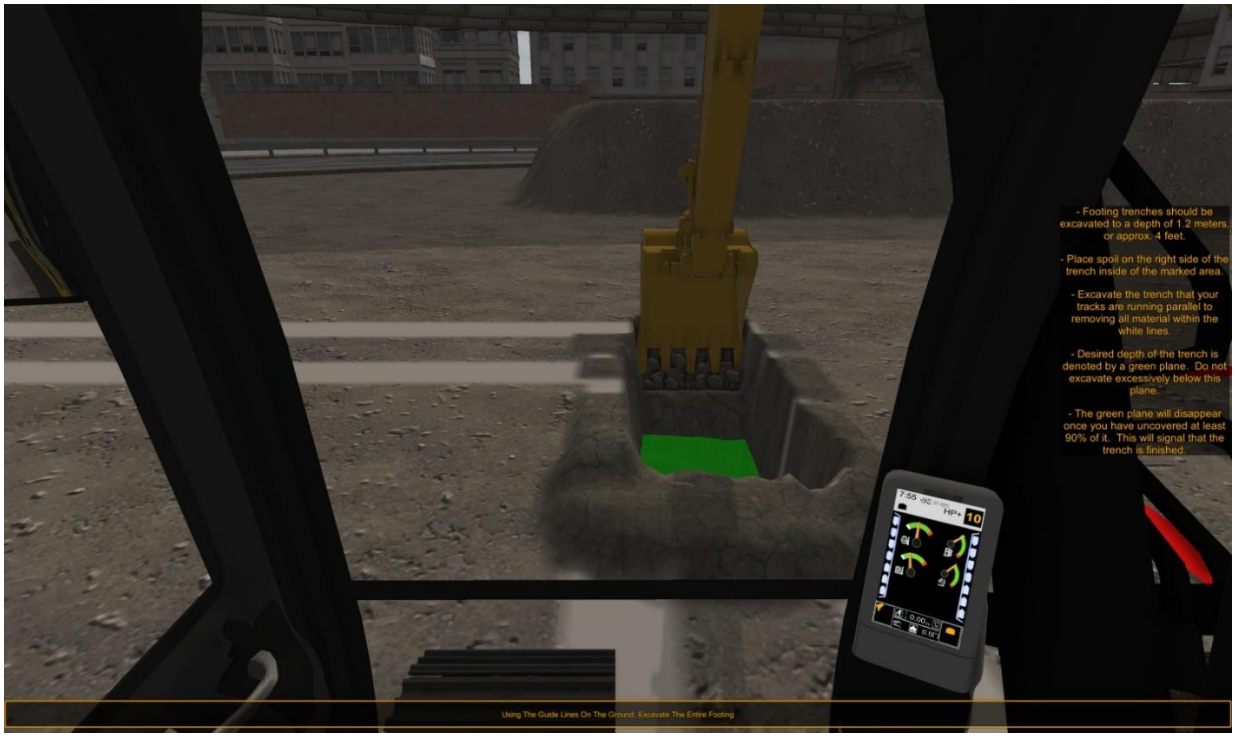


Figure 4-7. Collision Due to Premature Backtracking



Figure 4-8. Incorrect Dump Placement



Figure 4-9. Jerky Dumping Motion

5.0 Conclusion

This study developed and validated a multimodal, multi-agent AI framework for objective assessment of excavator operator proficiency through temporal behavior analysis. The framework operationalizes three complementary metrics: the Simultaneity Index, Bimanual Coordination Score, and multi-control usage distributions, that capture distinct dimensions of operator coordination difficult to quantify through conventional productivity measures. The analysis confirmed systematic differences in temporal coordination across skill levels. Expert operators demonstrated sustained simultaneous control usage and higher bimanual coordination, with the Arm–Boom pair serving as the strongest skill discriminator (novice: 0.125, expert: 0.542). Swing integration emerged as a late-developing marker of expertise rather than an incrementally acquired skill. The BCS analysis revealed not only significant mean differences across groups (novice: 0.109, intermediate: 0.157, expert: 0.333) but also distinct distributional patterns, with experts forming tight clusters while novices showed high variability. The framework demonstrated practical value through AI-generated training reports that produced qualitatively distinct behavioral interpretations for different skill levels, diagnosing specific coordination patterns such as "Stop-and-Go Swing" and "Sequential Isolation." The automated video analysis successfully detected operational irregularities, with independent expert review confirming alignment between automated detection and instructor-level evaluation. A key methodological strength is the use of industry-standard simulators actively deployed in professional training environments, ensuring ecological validity and direct transferability. The framework transforms raw control signals into interpretable behavioral insights and actionable training guidance, positioning it as a continuous, adaptive training partner rather than merely an assessment instrument. Future deployment as a real-time guidance system could provide live feedback during training sessions, supporting improved operational efficiency, safety, and skill development in construction equipment operation.

6.0 Reference

ABC. (2022, February 22). ABC: Construction Industry Faces Workforce Shortage of 650,000 in 2022. *ABC: Construction Industry Faces Workforce Shortage of 650,000 in 2022*. https://www.abc.org/News-Media/News-Releases/abc-construction-industry-faces-workforce-shortage-of-650-000-in-2022

AGC. (2024). *2024 Workforce Survey Analysis*. Associated General Contractors of America (AGC). https://www.agc.org/sites/default/files/Files/Communications/2024_Workforce_Survey_Analysis.pdf

Ahn, S., Kim, T., Park, Y.-J., & Kim, J.-M. (2020a). Improving Effectiveness of Safety Training at Construction Worksite Using 3D BIM Simulation. *Advances in Civil Engineering*, *2020*(1), 2473138. https://doi.org/10.1155/2020/2473138

Ahn, S., Kim, T., Park, Y.-J., & Kim, J.-M. (2020b). Improving Effectiveness of Safety Training at Construction Worksite Using 3D BIM Simulation. *Advances in Civil Engineering*, *2020*(1), 2473138. https://doi.org/10.1155/2020/2473138

Bae, J., Kim, K., & Hong, D. (2019). Automatic Identification of Excavator Activities Using Joystick Signals. *International Journal of Precision Engineering and Manufacturing*, *20*(12), 2101–2107. https://doi.org/10.1007/s12541-019-00219-5

Bernold, L. E. (2007). Quantitative Assessment of Backhoe Operator Skill. *Journal of Construction Engineering and Management*, *133*(11), 889–899. https://doi.org/10.1061/(ASCE)0733-9364(2007)133:11(889)

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., Arx, S. von, Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., Brynjolfsson, E., Buch, S., Card, D., Castellon, R., Chatterji, N., Chen, A., Creel, K., Davis, J. Q., Demszky, D., … Liang, P. (2022). *On the Opportunities and Risks of Foundation Models* (arXiv:2108.07258). arXiv. https://doi.org/10.48550/arXiv.2108.07258

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). *Language Models are Few-Shot Learners* (arXiv:2005.14165). arXiv. https://doi.org/10.48550/arXiv.2005.14165

CAT. (2025). *CAT Simulators*. CAT Simulators. https://catsimulators.com/construction-simulators/

Fan, W., Chen, P., Shi, D., Guo, X., & Kou, L. (2021a). Multi-agent modeling and simulation in the AI age. *Tsinghua Science and Technology*, *26*(5), 608–624. https://doi.org/10.26599/TST.2021.9010005

Fan, W., Chen, P., Shi, D., Guo, X., & Kou, L. (2021b). Multi-agent modeling and simulation in the AI age. *Tsinghua Science and Technology*, *26*(5), 608–624. https://doi.org/10.26599/TST.2021.9010005

Fletcher, S., & Islam, M. Z. (2018). Comparing sets of patterns with the Jaccard index. *Australasian Journal of Information Systems*, *22*. https://doi.org/10.3127/ajis.v22i0.1538

Fulford, R., & Standing, C. (2014). Construction industry productivity and the potential for collaborative practice. *International Journal of Project Management*, *32*(2), 315–326. https://doi.org/10.1016/j.ijproman.2013.05.007

Funke, I., Mees, S. T., Weitz, J., & Speidel, S. (2019). Video-based surgical skill assessment using 3D convolutional neural networks. *International Journal of Computer Assisted Radiology and Surgery*, *14*(7), 1217–1225. https://doi.org/10.1007/s11548-019-01995-1

Lee, J., & Lee, S. (2023). Construction Site Safety Management: A Computer Vision and Deep Learning Approach. *Sensors*, *23*(2), 944. https://doi.org/10.3390/s23020944

Lee, J. S., Ham, Y., Park, H., & Kim, J. (2022). Challenges, tasks, and opportunities in teleoperation of excavator toward human-in-the-loop construction automation. *Automation in Construction*, *135*, 104119. https://doi.org/10.1016/j.autcon.2021.104119

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2021). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks* (arXiv:2005.11401). arXiv. https://doi.org/10.48550/arXiv.2005.11401

Liu, Y., Zhou, Y., Wen, S., & Tang, C. (2014). A Strategy on Selecting Performance Metrics for Classifier Evaluation: *International Journal of Mobile Computing and Multimedia Communications*, *6*(4), 20–35. https://doi.org/10.4018/IJMCMC.2014100102

Luciano da F, C. (2021). *Further generalizations of the Jaccard index*. https://doi.org/arXiv%2520preprint%2520arXiv:2110.09619

Miyazaki, R., Akiyama, M., Koiwai, K., Kozui, M., Yamashita, K., & Yamamoto, T. (2021). Construction of a Skill Evaluation System Based on DATABASE-DRIVEN Approach for a Hydraulic Excavator. *IEEJ Transactions on Electrical and Electronic Engineering*, *16*(10), 1390–1396. https://doi.org/10.1002/tee.23435

Narayanan, S., & Georgiou, P. G. (2013). Behavioral Signal Processing: Deriving Human Behavioral Informatics From Speech and Language. *Proceedings of the IEEE*, *101*(5), 1203–1233. https://doi.org/10.1109/JPROC.2012.2236291

NCCER. (2023). *Heavy Equipment Operations Training Standards*. National Center for Construction Education and Research (NCCER).

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V.,

Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2024). *GPT-4 Technical Report* (arXiv:2303.08774). arXiv. https://doi.org/10.48550/arXiv.2303.08774

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). Generative Agents: Interactive Simulacra of Human Behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*, 1–22. https://doi.org/10.1145/3586183.3606763

Proctor, R., Thomas, B. S., & Dunston, P. (2022). Human Interactions with an Excavator Simulator: Assessing Two Patterns of Joystick Control Mappings. In *Human-Automation Interaction: Manufacturing, Services, and User Experience* (Vol. 10). Springer International Publishing.

Rueda-Delgado, L. M., Solesio-Jofre, E., Serrien, D. J., Mantini, D., Daffertshofer, A., & Swinnen, S. P. (2014). Understanding bimanual coordination across small time scales from an electrophysiological perspective. *Neuroscience & Biobehavioral Reviews*, *47*, 614–635. https://doi.org/10.1016/j.neubiorev.2014.10.003

Sekizuka, R., Ito, M., Saiki, S., Yamazaki, Y., & Kurita, Y. (2020). System to Evaluate the Skill of Operating Hydraulic Excavators Using a Remote Controlled Excavator and Virtual Reality. *Frontiers in Robotics and AI*, *6*, 142. https://doi.org/10.3389/frobt.2019.00142

Slutski, L., Gurevich, I., & Edan, Y. (2000). Analysis and enhancement of human-machine interfaces using a joystick controller. *Human Factors and Ergonomics in Manufacturing*, *10*(2), 161–175. https://doi.org/10.1002/(SICI)1520-6564(200021)10:2%253C161::AID-HFM3%253E3.0.CO;2-W

Song, H., Kim, T., Kim, J., Ahn, D., & Kang, Y. (2021). Effectiveness of VR crane training with head-mounted display: Double mediation of presence and perceived usefulness. *Automation in Construction*, *122*, 103506. https://doi.org/10.1016/j.autcon.2020.103506

Taheri Gorji, H., Wilson, N., VanBree, J., Hoffmann, B., Petros, T., & Tavakolian, K. (2023). Using machine learning methods and EEG to discriminate aircraft pilot cognitive workload during flight. *Scientific Reports*, *13*(1), 2507. https://doi.org/10.1038/s41598-023-29647-0

Team, G., Georgiev, P., Lei, V. I., Burnell, R., Bai, L., Gulati, A., Tanzer, G., Vincent, D., Pan, Z., Wang, S., Mariooryad, S., Ding, Y., Geng, X., Alcober, F., Frostig, R., Omernick, M., Walker, L., Paduraru, C., Sorokin, C., … Vinyals, O. (2024). *Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context* (arXiv:2403.05530). arXiv. https://doi.org/10.48550/arXiv.2403.05530

Tichon, J., & Phil, D. (2010). Plant operator simulation: Benefits and drawbacks for a construction training organization. *Cognition, Technology & Work*, *3*, 219–229.

Vahdatikhaki, F., Langroodi, A. K., Scholtenhuis, L. O., & Dorée, A. (2022). Feedback support system for training of excavator operators. *Automation in Construction*, *136*, 104188. https://doi.org/10.1016/j.autcon.2022.104188

Vorontsov, I. E., Kulakovskiy, I. V., & Makeev, V. J. (2013). Jaccard index based similarity measure to compare transcription factor binding site models. *Algorithms for Molecular Biology*, *8*(1), 23. https://doi.org/10.1186/1748-7188-8-23

Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., Jiang, L., Zhang, X., Zhang, S., Liu, J., Awadallah, A., White, R. W., Burger, D., & Wang, C. (n.d.-a). *AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation*.

Wu, Q., Bansal, G., Zhang, J., Wu, Y., Li, B., Zhu, E., Jiang, L., Zhang, X., Zhang, S., Liu, J., Awadallah, A., White, R. W., Burger, D., & Wang, C. (n.d.-b). *AutoGen: Enabling Next-Gen LLM Applications via Multi-Agent Conversation*.

Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., Zheng, R., Fan, X., Wang, X., Xiong, L., Zhou, Y., Wang, W., Jiang, C., Zou, Y., Liu, X., … Gui, T. (2023). *The Rise and Potential of Large Language Model Based Agents: A Survey* (arXiv:2309.07864). arXiv. https://doi.org/10.48550/arXiv.2309.07864