Final Report

Behavioural Change Assessment for AI Monitoring of animal welfare

Project code
2026-1031

Prepared by
Impetus Animal Welfare Ltd
Dr Michael Patching

Published by
Australian Meat Processor Corporation

Date submitted
07/04/2026

Contents

1.0 Abstract

This project evaluated the monitoring utility and organisational conditions present when using AI-based livestock monitoring systems and if these lead to measurable improvements in animal welfare in commercial beef processing facilities. While these systems can detect welfare related events continuously and objectively, their impact depends on how the data is used within the business.

Using AI monitoring data from a commercial facility, supported by staff feedback and Welfare Footprint Framework (WFF) analysis, the project established a baseline of animal handling practices and identified key patterns in behaviour. The results showed that AI monitoring provides a more accurate and complete picture of handling practices than traditional observation, but does not by itself drive improvement.

The main finding is that management systems, not technology alone, determine whether welfare outcomes improve. Where monitoring data is actively reviewed, linked to corrective actions, and used to inform operational and infrastructure decisions, it has strong potential to improve welfare. Where this governance is absent, the system functions as a measurement tool without sustained change.

The project also demonstrated how WFF can be used to quantify welfare outcomes and compare the impact of different interventions. This provides processors with a novel practical method to measure and internally report welfare improvements in a consistent and auditable way.

The report provides guidance for the red meat processing industry on how to implement AI monitoring systems effectively, highlighting the importance of management accountability, structured review processes, and integration with operational decision making.

2.0 Executive summary

The Australian red meat processing industry has begun investing in AI-based livestock welfare monitoring. These systems are technically capable of detecting welfare relevant events. But the critical question of whether this monitoring capability translates into actual behavioural change and improved welfare outcomes for animals, depends far less on the technology itself than on the management structures surrounding it.

This project examined that question at an Australian cattle abattoir using the Argus AI system. The central finding is that the QA team engaged meaningfully with the system and used it productively, and that further structured management pathways from detections and data to corrective actions and facility decisions would allow the full potential of the monitoring investment. This reflects a pattern likely to be common across the industry as AI monitoring systems are deployed. Governance structures need to be built alongside the deployment of technology.

2.1 Key Findings

- AI monitoring can provide more complete and consistent data than intermittent observation. Staff and management at facilities commonly hold beliefs about when and where welfare events occur (assumptions about time of day, stock type, or operator) that are not (necessarily) reliably accurate. AI data covering the full operational day consistently reveals patterns that anecdotal observation can miss or distort.

- Footage review is essential and cannot be replaced by data alone, including for advanced vision systems. AI systems flag events and generate statistics, but they cannot interpret context. In the complex, dynamic environment of a livestock processing facility, human review of flagged footage is what converts a detection into an understanding of the real issue of whether it is an infrastructure problem, a training gap, a particular operator, or a stock characteristic. Metrics without footage review lead to misdiagnosis and ineffective interventions.
- Management engagement is the primary determinant of whether AI monitoring drives change. Where detections are reviewed by QA but findings are not systematically converted into corrective actions for the livestock team, and where infrastructure or design issues identified through monitoring are not escalated to capital planning, the system functions as a measurement tool only. This may still have value for the business, especially for KPI reporting. However for sustained welfare improvement, and risk reduction, there is a requirement for management ownership of the findings, not just QA ownership of the system.
- Electric goad use at the facility illustrated the limits of responses that were training only. Despite sustained monitoring and periods of staff training, goad use was still required and did not show a durable reduction. Analysis indicated that in several sections of the race, goad use is driven by facility geometry and animal flow constraints rather than operator attitude or capability. Logically, training interventions applied to an infrastructure problem produce limited and unsustained change. The data from this project makes a clear case for a structured facility design review in those areas.
- The staff perception survey (n = 4) is indicative only but reveals a meaningful perception gap. QA staff and management perceived the usefulness of the systems differently which is a predictable adoption risk that communication and system accuracy improvements must address.
- The Welfare Footprint Framework provides a quantitative welfare baseline that enables processors to demonstrate the impact of interventions. WFF analysis used the AI event data to quantify welfare burden that could be used to compare outcomes before and after targeted changes.

2.2 Benefits to Industry

This project provides the red meat processing industry with an evidence-based framework for realising the full value of AI welfare monitoring investment. The combination of AI event monitoring, structured management accountability, footage review protocols, and WFF-based welfare burden measurement creates a continuous improvement system that can be adopted and scaled across processing sites. The industry guidance in Section 9 of this Report provides practical implementation steps.

3.0 Introduction

The deployment of AI-based livestock monitoring systems in Australian abattoirs represents a technological advance for animal welfare assurance. These systems can offer continuous, objective, time stamped records of handling events across the full operational day which has not previously been available to the industry. The systems typically provide a higher level of monitoring than current systems that use either periodic samples or point-in-time audit snapshots.

Yet the question of whether this capability translates into better welfare outcomes for animals is not a technology question. It is a management question. Monitoring data that is reviewed by a QA team but not connected to corrective action systems for the livestock team, infrastructure investment decisions, or operational management accountability will not change how animals are handled. The technology maintains oversight of a large range of events, 24 hours per day. Whether anything changes depends on what organisations do with the information that technology reports to them.

This project was designed to test that proposition at a cattle abattoir. Argus AI had been operational at the facility for approximately one year before the formal project period began. Staff had been exposed to monitoring for an extended period. Some of the QA team had engaged with the system consistently.

A secondary but important contribution of this project is the application of the Welfare Footprint Framework (WFF; Alonso & Schuck-Paim, 2025) in a commercial processing environment. WFF provides a method for expressing welfare outcomes in the time-based units of welfare burden minutes per animal, that are especially useful for internal reporting on the impact of change. The analysis conducted in this project establishes a baseline welfare burden structure that processors or the industry can use to demonstrate the quantitative welfare impact of specific interventions over time.

Critically, this project also reinforces the irreplaceable role of human footage review alongside AI-generated data. AI systems classify events and create structured data, they do not and at this stage should not interpret them. Understanding whether an elevated goad detection reflects a training problem, an infrastructure constraint, a particular stock class, or a specific operator requires a human reviewer watching the footage at the timestamp of the event. The industry's adoption of AI monitoring will only deliver its full potential if the operational culture and time allocation for this footage review is established and protected.

4.0 Project objectives

4.1 Achievement Against Project Objectives

The table below summarises performance against the objectives defined in AMPC Research Agreement 2026-1031 (Variation Agreement #1), based on the evidence generated in this project.

Project Objective	Outcome	Supporting Evidence
Benchmark key animal handling behaviours using Welfare Footprint Framework-aligned metrics	A quantified welfare burden baseline was established across four processing modules (receival, lairage, race, stunning/slaughter) using WFF methodology. The figures published here use the illustrative reference dataset in Sections 8.1–8.2; facility-specific values are held in the confidential annex	Sections 8.1 and 8.2
Quantify changes in staff behaviour and welfare outcomes following AI-assisted monitoring implementation	Early phase behavioural patterns and trends were identified using retrospective and active monitoring data; however, a full longitudinal pre/post comparison was constrained by the available dataset	Section 6.0 Results Part A; confidential data annex
Validate the effectiveness of Argus AI as a driver for	Argus AI demonstrated operational utility as a detection and monitoring system. Results show that measurable welfare improvement depends	Sections 6.1–6.4, 7.0 and 9.0

measurable welfare improvements	on management systems rather than technology alone	
Demonstrate the feasibility of integrating WFF methodology into routine operational data evaluation	WFF was applied to indicative data that was supported by AI-derived event data, demonstrating a practical method for converting operational data into quantified welfare outcomes	Section 8.0 Results Part C
Evaluate staff performance outcomes related to animal welfare post-AI implementation to determine value to industry	Staff perception survey identified key themes including workload impacts, perception gaps, and barriers to adoption; results are indicative due to small sample size (n=4)	Section 7.0 Results Part B
Deliver industry guidance and a roadmap for sector-wide adoption of AI-enhanced welfare monitoring	Practical guidance on governance, implementation, and system integration was developed and presented for both facility-level and industry-level adoption	Sections 9.0 Discussion and 11.0 Recommendations

Overall, the project successfully delivered against its primary objectives, with the key finding that the effectiveness of AI monitoring in improving animal welfare outcomes is determined by the management systems and governance structures into which it is integrated.

5.0 Methodology

The project used a mixed-methods retrospective data review and early-phase evaluation, comprising planning, AI monitoring and final evaluation. Monitoring and data collection took place from August-November 2025 (as the system was functional during this period). The methodology comprised three distinct but complementary analytical components, each addressed separately in the Results sections below.

5.1 AI Monitoring and Behavioural Change Analysis

The Argus AI system was fully operational at the facility from early 2025. Due to the nature of the project operating in a commercial facility with existing use of the system and data being available, this event data was incorporated into the project with agreement of all parties. This provided a cleaner dataset than the later project period, during which engagement with the system was reduced, though it introduces the methodological caveat that it was difficult to determine a baseline before intervention. There was also no significant or specific animal welfare compliance gap at the facility that needed to be addressed, and so behaviour change was to achieve more optimal outcomes rather than addressing a compliance deficit.

A dedicated facility system operator was appointed to validate daily AI detections, classify true positive events, remove false-positives, and add operational context notes. Fortnightly review meetings involving QA personnel, livestock management, plant management, and Impetus project management provided the forum for translating detections into operational findings. Key event categories tracked included electric goad use (total and excessive), precautionary and ineffective re-stuns, captive bolt use, down animals, and rough handling events, as well as counting cattle unloading off the truck.

The analytical workflow assessed event trends over time, patterns by time of day and day-of-week, relationships between event types, and the degree to which detections were reviewed, validated, and converted into corrective actions. This last dimension, the conversion rate from detection to action, is the primary indicator of management engagement with the system.

5.2 Staff Perception Survey

A structured online survey (see Appendix 2) was developed and distributed to staff with direct experience of AI monitoring systems in use across the participating business. The survey assessed perceived usefulness and accuracy, workload impacts, attitudes toward monitoring (supportive versus punitive), confidence in the technology, and barriers to adoption and suggestions for improvement. The survey was designed to capture both the QA team's operational experience and management's strategic perspective.

Four responses were received. This is a small sample size and results are treated as indicative of qualitative themes rather than statistically representative findings. This limitation is explicitly acknowledged throughout.

5.3 Welfare Footprint Framework Analysis

The Welfare Footprint Framework (WFF; Alonso & Schuck-Paim, 2025) expresses welfare impact as the cumulative time an animal spends in negative affective states of defined intensity. For each harm the experience is decomposed into phases (a Pain-Track; Alonso & Schuck-Paim, 2021), the probability of each intensity category is estimated per phase from evidence, and phase duration is multiplied by those probabilities to give Cumulative Pain (the time spent in each intensity category) which is then scaled by the proportion of animals affected.

Four intensity categories are defined by how far the experience disrupts normal function. Annoying (little disruption), Hurtful (partial disruption, motivated behaviour declines), Disabling (major disruption, normal behaviour largely prevented, potent analgesia required) and Excruciating (complete disruption, overriding all else). The categories are reported separately rather than collapsed into a single weighted score, and all durations are expressed in hours with sleep and other non-conscious time excluded.

The WFF is an analytical framework that can incorporate measured event and timing data. It converts measured event and timing data into a welfare figure, which means its accuracy and its usefulness depend on the data a facility collects. The analysis below makes explicit, harm by harm, which data drive each result.

6.0 Results Part A: AI Monitoring and Behavioural Change

The results presented in this section are based on a retrospective and early-phase monitoring dataset (August–November 2025), during which the Argus AI system was actively used and reviewed within the facility. This period provided the most reliable dataset for analysis, as it reflects consistent system engagement, validation of detections, and regular operational review.

While the original project design anticipated a longer term longitudinal comparison, the available dataset still enables characterisation, identification of behavioural patterns, and evaluation of early stage responses to AI-supported monitoring. These findings provide a valid foundation for assessing both system capability and the organisational conditions required for sustained behavioural change.

6.1 The Value of AI Data Over Anecdotal Observation

One of the most practically significant findings of this project is the contrast between what operational staff believed to be happening and what the AI data showed. Prior to systematic AI monitoring, facility staff and management held a range of assumptions about handling patterns. Beliefs included which times of day saw the most goad use, which operators were responsible for most incidents, and what the typical ratio of precautionary to ineffective stuns was. In almost every case, the AI data provided a more complete and often different picture.

> *The full-day, continuous nature of AI data is its primary advantage over human observation and anecdote. A QA officer observing a race during a spot check sees minutes out of hours. AI monitoring sees every event, every shift, every day and can show patterns that are invisible to intermittent human observation.*

Two specific examples illustrate this:

Electric Goad Use (Day-of-Week Pattern)

Case study analysis of goad use in October 2025 identified a consistent pattern of relatively elevated goad events toward the end of the working week. This pattern, not previously identified through anecdotal observation or QA spot checks, suggests a combination of factors that may include stock class variation across the week and potential accumulation of staff fatigue. Without AI data spanning the full week, this pattern would remain invisible and therefore unaddressed. In businesses, these results could allow for more targeted monitoring of individual use or to link stock lines to outcomes.

Precautionary vs. Ineffective Stuns

The AI data provided the first reliable breakdown of re-stun events by type, distinguishing precautionary stuns (operator initiated as a precaution following a standard stun) from genuinely ineffective stuns (stun operator’s own assessment of failed effectiveness). Precautionary stuns showed a slight reduction trend from October to November 2025, while genuine ineffective stuns remained low. This distinction matters as the cause and impact of a high precautionary stun rate is different from the response to a high ineffective stun rate (potential for conscious cattle after a stun). Anecdotal reporting or vision-only systems have not yet been developed to reliably make this distinction and so video review or input from staff on the floor remains important. Argus utilises a button at the restraint device that the staff can push to log in ineffective stuns, which then automatically correlate to the event.

Indicator	Baseline Trend (Aug–Nov 2025)	Management Action Required
Electric goad use (total)	Consistent but variable through week. No sustained reduction	Infrastructure review of high-use race sections. Staffing review for late week staff schedules or training.
Excessive goad use	Directly mirrors total goad use trend	Define an escalation threshold and differentiate infrastructure from training drivers by location.

Precautionary re-stuns	Slight reduction Oct–Nov 2025	Monitor trends. Align this with stunner maintenance schedule.
Ineffective stuns	Low. Consistent	Continue monitoring. No immediate intervention indicated.
Captive bolt use	Variable. Spikes identify maintenance/anomaly events	Use as an indicator of risk or operational issues; footage review is important.
Total detections	Increasing as system matures and has the capability of detecting more	Increasing detections may reflect system calibration improvement, not deteriorating welfare.
Verified findings vs. detections	Slight decrease in total findings despite increasing detections	Finding rate is driven by time allocated to system review which is a management resourcing question.

6.2 Alignment of all Parties

The QA team engaged productively with the Argus system and reviewed detections, validated events, and flagged issues. The regular meetings identified the importance of all parties to optimise the outcomes from the data collection through a systems approach between different parts of the business.

Data without action is measurement, not improvement. The gap between QA owning the monitoring system and management owning the response to what the system finds is where welfare improvement potential can be lost.

This gap manifested in several ways:

- Livestock team corrective actions: when goad use was identified as elevated in a specific location on specific days, the pathway to briefing the livestock team, implementing a targeted intervention, and verifying its effect was inconsistent. The monitoring data existed but the governance structure to act on it did not.
- Infrastructure and design escalation: analysis of goad-use patterns indicated that some sections of the race may constrain animal flow and warrant further facility review, independently of operator attitude or training.
- Spikes in re-stun rates: analysis identified instances in which re-stun rates were elevated above the normal operating range. The gap identified here was regarding how corrective actions were reported into a system for use during audits.

It is important to note that this finding reflects a systems design issue that is likely to be common across the industry and not of any individual or team at the facility. Most Australian facilities have built their animal welfare management systems around audits, anecdotal observation, and compliance reporting, and not continuous data streams. If AI monitoring systems are deployed into these facilities it is likely that the AI will augment data collection, and not an active management response. This will be more pronounced in businesses that have undergone limited or incomplete digital transformation or with resource constraints.

6.3 Electric Goad Use: The Infrastructure Ceiling

Electric goad use is the clearest example in this dataset of the difference between a training problem and an infrastructure problem, and the risk of applying training-based interventions only. During the reporting period the business was working towards elimination of goad use.

The AI data enabled location-specific analysis, identifying that some devices were associated with higher use rates than others and that certain physical sections of the race consistently generated more goad events. Staff training was applied at periods during the project, and short term reductions were quantified following training sessions, but these reductions were not sustained.

Discussion with staff and management, informed by review of footage at flagged timestamps, indicated that cattle movement may have been constrained by geometry or flow design, increasing the likelihood of goad use. Training equips operators with the knowledge that goad use should be minimised however it cannot change the physical environment that makes that minimisation difficult.

The AI monitoring data's value here is to help identify the problem (location-specific, time-of-day-specific, device-specific), and document the impact of change following training, or to make the case for a different type of intervention. It is challenging to have the same level of confidence from anecdotal observation or periodic audits.

6.4 The Importance of Footage Review

Throughout the project, the most operationally useful findings came not from analysis of aggregate detection data but from human review of footage at the timestamps of specific events. AI systems classify events, they cannot interpret them. Two events that generate the same detection classification can have entirely different causes and require entirely different responses.

Examples from this project illustrate the point:

- A spike in back-up captive bolt use with a review of footage at those timestamps showed the events related to euthanasia of a compromised animal on the edge of the camera view. Without footage review it is unlikely that vision systems alone would have detected and reported on the issue as it was out of main view. Pulling the back-up stunner became an important proxy indicator of an event that needed review.
- Excessive goad use detections in a specific race section, when reviewed on footage, showed cattle hesitating at a specific point in the race in a pattern consistent with a visual obstruction or lighting issue, not an operator attitude problem. This insight directly informed an infrastructure discussion.
- Captive bolt stun events reviewed on footage allowed differentiation between stuns performed on the cradle and stuns performed in the box, a distinction that matters for both welfare interpretation and equipment attribution that the detection data alone could not provide.

As AI systems improve, the temptation will be to rely on the data without reviewing the footage. This project strongly cautions against that tendency. The complexity of livestock processing environments with variable animal behaviour, equipment states, operator contexts, and facility conditions means that human review of flagged footage remains essential for accurate interpretation. Data tells you something happened. Footage tells you what actually happened.

Results Part B: Staff Attitudes Survey

The survey returned only four responses. Results are indicative of qualitative themes and perception patterns only. No statistically robust conclusions can be drawn from the small dataset. The value of this data is in the direction and nature of the perceptions it reveals, not in any claim to representativeness.

7.1 Perceived Value and Accuracy

Because the respondent group is small, results are reported as themes rather than as individual or role-level ratings. All respondents agreed or strongly agreed that the system could improve animal welfare outcomes and staff training, and all rated its potential value to the business positively. Ratings of the system's current accuracy and of confidence that a flagged event was worth reviewing were consistently lower among respondents working with the system daily than among those using its outputs for reporting and oversight. Respondents closest to the system reported increased review time, while respondents further from it reported improved information quality and workflow.

The direction of this gap matters more than its magnitude. It reflects the operational cost of managing false-positive detections, which falls on the reviewing function rather than on those receiving the reports. This is a consistent pattern in technology adoption, where those closest to the operational burden of managing system imperfections experience least of the strategic benefit.

7.2 Attitudes Toward Monitoring Purpose

Responses are again reported thematically. All respondents agreed that the system helps identify real operational risks, could improve animal welfare outcomes, helps resolve disagreements about incidents, and could improve the training of new staff. Two areas showed divergence. First, whether the system adds extra workload: respondents operating it daily agreed that it does, while management respondents did not. Second, and more significantly, whether the system is mainly used for oversight of staff: this was agreed by the operational respondent and disagreed with by both management respondents.

The divergence on monitoring purpose is the more important of the two. If operational staff believe they are under surveillance rather than supported, their engagement with the system and their willingness to act on its findings will be compromised. This is a communication and culture issue and it sits with management to address. It is also a predictable one, and any facility deploying AI monitoring should expect to have to manage it explicitly rather than assume it away.

7.3 Barriers to Adoption

Respondents were asked to identify barriers to effective system use. Combined responses identified:

- False positive detections consuming disproportionate QA time (cited by 3 of 4 respondents). This was the single most consistent barrier and is a major limitation of vision-only systems in complex environments
- Insufficient time during operational shifts for systematic review
- Management not prioritising system engagement as a clear expectation
- Staff turnover resulting in loss of institutional knowledge about system use
- Unclear ownership of who is responsible for responding to detections

7.4 Long-Term Sustainability

Responses diverged on whether the system would still be in use in 12 months. Management responses were more confident than operational responses, and the concerns cited by operational respondents were inaccurate KPIs and time consumed by false-positive detections. This divergence is a significant finding in the survey. A system perceived by the people who operate it daily as a net burden is at risk of quiet abandonment regardless of management intent. Reducing false-positive rates and managing expectations, while also demonstrating that system findings lead to tangible operational improvements are the most important factors in reversing this perception.

Vision-based monitoring performance is sensitive to camera position, field of view, lighting, occlusion, animal variation and the number of event classes being assessed. Procurement decisions should therefore be based on site-specific validation of accuracy, false-positive workload, coverage, computing requirements and total cost of operation. This project did not undertake a comparative evaluation of competing commercial platforms.

Results Part C: Welfare Footprint Framework Analysis

A central result of this project is that the AI monitoring outputs (timestamps, goad counts, re-stun classifications and event frequencies) can then be taken through the WFF methodology to produce a quantified output. The Framework works best where there are actual data inputs to inform the impact of interventions on groups of animals or company wide.

8.1 The six pre-slaughter welfare harms

Each harm below carries a short statement of the affective experience and its intensity and duration profile, the factors that materially affect the outcome (the levers a facility can pull), and the data that measures it. These are the reference points the Discussion (Section 9.5) draws on when weighing where welfare investment delivers a return.

Extended holding (Hunger)

A standard overnight animal (around 24 hours off feed) accrues roughly 11.5 hours of Annoying and 6.8 hours of Hurtful experience.

Progressive feed deprivation across transport and lairage produces hunger that escalates with cumulative time off feed. However the ceiling is low and even a 72-hour fast may not be a major physiological stressor (van der Walt et al., 1993) and cattle tolerate deprivation to 48 hours with minimal behavioural change (Bond et al., 1976). From a WFF perspective this means that the experience is time dependent and Annoying-dominated, becoming Hurtful over hours, with a small Disabling tail only at the 48-hour extreme. Australian data corroborate the mechanism directly as liver glycogen falls monotonically with lairage duration across short medium and long lairage (George & George, 2023).

Factors that materially affect it: cumulative time off feed (the dominant lever), transport duration and on-farm curfew before arrival, and the point at which lairage feed is provided.

Data that measures it: arrival and slaughter timestamps (off-feed duration).

Lairage (social and environmental disruption)

A standard-overnight animal accrues roughly 5.8 hours of Annoying and 1.7 hours of Hurtful experience.

In this facility the WFF reviewed the impact of fear and distress of an unfamiliar lairage environment, with the cohort kept intact so that no within group regrouping occurs. In cattle the acute novelty response resolves quickly and cortisol returns to baseline within about 70 minutes (Veissier & Le Neindre, 1988) and this is buffered further by familiar penmates (Mazer et al., 2020; Kiyokawa et al., 2014). However cattle are likely to never fully adapt within the holding period and rest poorly (Cockram, 1991), and lying deprivation is genuinely aversive (Munksgaard & Simonsen, 1996), leaving a likely persistent low-grade background for most animals. The profile is therefore front-loaded and then flat with a fixed Hurtful spike in the first hours followed by a long Annoying tail, with no Disabling and no Excruciating, because the fear of a functional, cohort-buffered animal does not meet the Disabling threshold.

Factors that materially affect it: group mixing (avoided here, a major reduction), stocking density, flooring and lying comfort, and total lairage time.

Data that measures it: lairage arrival-to-exit timestamps, batch mixing records, and pen stocking density.

Thermal discomfort (HLI)

On a normal hot-season day a non-shaded Bos taurus animal accrues roughly 2.9 hours of Annoying, 2.0 hours of Hurtful and 0.45 hours of Disabling experience per hot-day exposure.

Heat load, quantified by the Heat Load Index (HLI; Gaughan et al., 2008) including black globe temperature, humidity and wind, with Accumulated Heat Load Units for the duration component. HLI is associated with panting score and can be used to map to WFF intensity, with slight panting Annoying, fast panting is Hurtful, and open-mouth panting with feeding and rumination ceasing is Disabling. Shade can reduce heat stress depending on weather, stocking and facility conditions and we have selected this to use in the illustrative model (Sullivan et al., 2011; Gaughan et al., 2010; Lees et al., 2020; (Schuck-Paim et al., 2026).

Factors that materially affect it: genotype (Bos indicus threshold 96 versus Bos taurus 86, which largely removes the harm on normal hot days; Gaughan et al., 2010), shade (which raises the HLI threshold by 3 to 7 and roughly halves panting), stocking density, and season.

Data that measures it: HLI inputs (black globe temperature, humidity, wind), panting score, a hot-season flag, and genotype.

Ineffective stun (return to sensibility)

A genuine complete stun failure with a fully sensible animal (not the common partial stun, which is re-stunned before sensibility returns), or consciousness at sticking. It is rare and genuine return-to-sensibility is around 0.16 per cent in fed steers and heifers (Grandin, 2002). This is against a repeat stun rate that is performed largely to reduce risk of poor stun outcomes, especially in non penetrative captive bolts and larger cattle. In cases of genuine stun failure, the impact per affected animal it is the only route to Excruciating, roughly 15 to 90 seconds of overwhelming distress, depending on whether the animal is promptly re-stunned, conscious on the cradle, or stuck without re-stun. At around 0.1 per cent prevalence this is a rounding error in accumulated time (about 0.01 seconds per average animal) yet remains a non-negotiable ethical base for all businesses.

Factors that materially affect it: stunner maintenance and cartridge condition, operator accuracy, monitoring and prompt re-stun, animal class (bulls and cows with heavy skulls carry higher rates; Grandin, 2002), and facility-type audit infrastructure.

Data that measures it: confirmed-sensibility re-stun rate (distinguished from precautionary re-stuns), stun-to-stick interval, and sensibility checks on the bleed rail.

Electric goad

An acute electric-shock episode, behaviourally very aversive (among the most aversive handling treatments; Pajor et al., 2000; Gellatly et al., 2019) but brief and non-tissue-damaging, with muted physiology for isolated shocks (Reinemann et al., 2003; Lefcourt et al., 1986; no substance-P rise, Tschoner et al., 2026). A single application is a roughly six-second Hurtful event with a Disabling instant. Repeated or excessive goading is different. Six prods over 5 to 10 minutes raised plasma lactate at slaughter (7.12 versus 4.29 millimoles per litre), a sustained sympatho-adrenal response, and produced measurably tougher, lower-grade meat in Australian feedlot cattle (Warner et al., 2007). Excessive use therefore adds a sustained fear-arousal overlay across the prodding episode, and is where the burden and the reducible loss concentrate and it is also why it is important to separately identify repeat electric goad use on one animal from general use.

Factors that materially affect it: goad prevalence and, especially, the excessive-use (three or more applications) sub-share, together with race geometry, lighting and flooring that cause animals to baulk, and handler training.

Data that measures it: goad-sensor counts and event timestamps, directly captured by the AI system, and the parameter most amenable to before-and-after intervention comparison.

Handling and restraint

The general fear of forced movement through the crowd pen and race, and of restraint in the stun box, for a normally-handled animal (the goad shock and the stun failure are counted separately). It is a brief, acute Annoying-to-Hurtful fear of around three minutes per animal, strongly dependent on facility design and temperament (Grandin & Shivley, 2015). Restraint type is the key lever: enforced head restraint roughly doubles cortisol at stick, from 67.6 to 143.1 nanomoles per litre (Ewbank et al., 1992), while good design lets cattle enter willingly and be stunned within about 10 seconds (Grandin, 1992). There is no Excruciating, and only a small Disabling possibility for the baulking minority at peak restraint.

Factors that materially affect it: race and crowd-pen design (curved race, lighting, non-slip floor, low-stress handling), restraint type and adjustment, and temperament and prior handling.

Data that measures it: vocalisation rate (the three per cent benchmark; Grandin, 2000), flight speed, race transit time, and baulking or electric-prod events.

8.2 Whole-of-process Cumulative Pain

Aggregating the six harms, with each population weighted by its prevalence and duration inputs. This gives the whole-of-process burden per average animal.

	Annoying	Hurtful	Disabling	Total
Baseline (non-shaded lairage)	16.2	7.8	0.12 (Excruciating ~0.00)	~24.1 h
With lairage shade	16.2	7.7	0.05 (Excruciating ~0.00)	~23.9 h

Table 1. Framework-derived Cumulative Pain, per average animal through the pre-slaughter process (hours). Excruciating experience is about 0.01 seconds per average animal and arises only from the ineffective stun. Indicative inputs; facility data replace them. Strictly additive (unadjusted for concurrent attention effects).

The contribution of each harm to the per-average-animal total is set out below.

Harm	Annoying	Hurtful	Disabling	Excruciating
Standard holding (hunger)	11.5	6.8	—	—
Lairage disruption	5.57	1.70	—	—
Thermal (non-shaded, hot-season share)	0.44	0.29	0.068	—
Ineffective stun	~0	~0	~0	0.000003
Electric goad	0.001	0.001	~0	—
Handling and restraint	0.030	0.023	~0	—

Table 2. Contribution by harm, per average animal (hours).

Lairage (hunger plus environmental disruption) accounts for approximately 96 per cent of the whole-of-process welfare burden, sitting almost entirely in the Annoying and Hurtful tiers. It is a duration effect of moderate discomfort endured for many hours by every animal. Disabling experience is small (about 0.12 hours) and almost entirely thermal, which is why shade is the clearest single lever on the severe tier. Excruciating experience is confined to the ineffective stun and is negligible in accumulated time, yet it is the sole ethical-ceiling harm.

8.3 Worked intervention examples

The framework's practical value is that a change to any single factor can be modelled and, once implemented, re-measured from the same data, with the welfare impact expressed in the same units. Three examples illustrate, each tied to a directly measurable input.

Lairage shade (thermal)

Providing effective lairage shade raises the HLI threshold by 3 to 7 and roughly halves panting on a hot day. Per hot-day exposure, per affected animal:

	Annoying	Hurtful	Disabling	Total
Non-shaded	2.93	1.95	0.45	5.3 h
Shaded	3.26	0.87	0	4.1 h

Shade eliminates the Disabling component entirely (0.45 hours to zero) and more than halves the Hurtful; the gain is concentrated in the severe tiers even though total time falls only modestly. This is verified before and after from HLI inputs and panting score (Sullivan et al., 2011; Gaughan et al., 2010; Lees et al., 2020).

Electric goad reduction (25 to 10 per cent of animals goaded)

Reducing the share of animals goaded, through race geometry, lighting, device-access control and handler training, cuts the goad-attributable burden by roughly 60 per cent:

Scenario	Goad-attributable time (per average animal)	Per 1,000 head	of which Disabling
25 per cent goaded (baseline)	~8.7 s	2.4 h	0.24 s
10 per cent goaded	~3.5 s	1.0 h	0.10 s
Saving	~5.2 s (about 60 per cent)	1.4 h	0.14 s

The reduction removes the sustained fear-arousal overlay of repeat prodding (the excessive-use sub-share, which carries the Warner et al. (2007) sympatho-adrenal response) together with the Disabling shock instants. At whole-of-process scale it barely moves the roughly 24-hour total as goad is well under 0.01 per cent of the burden, which is the audit paradox expressed in numbers, but it is the most directly measurable parameter (AI goad-sensor counts), so it is the easiest intervention to verify with before-and-after data, and it also removes the meat-quality penalty of excessive prodding.

Handling and restraint (poor versus good)

Handling burden is driven by facility design, restraint type and temperament. The table contrasts a poor scenario (a baulking race with distractions, rough handling, enforced head restraint that roughly doubles cortisol per Ewbank et al. (1992), and a fearful or extensively-reared animal) with a good scenario (a curved race with lighting and non-slip floor, low-stress handling, willing restraint with stunning within about 10 seconds, and a calm, habituated animal):

Scenario	Annoying	Hurtful	Disabling	Per animal	Per 1,000 head
Good handling	0.022	0.006	—	~1.7 min	~28 h
Poor handling	0.026	0.051	0.016	~5.6 min	~93 h

Good handling not only cuts total time in negative experience by around 3.3 times but shifts it out of the severe tiers as it is almost entirely Annoying, whereas poor handling produces substantial Hurtful and introduces Disabling fear reaching the activities-prevented level when a fearful animal is held in enforced restraint. Temperament is a genuine lever here (Grandin & Shivley, 2015): the same facility yields a calmer profile for habituated cattle, and acclimating animals to handling is itself an intervention. All of this is measurable from vocalisation rate, flight speed, race transit time and baulking events which were mostly scored with the AI system in place.

8.4 How the framework works in practice

The value of the framework is that it is populated by data a facility already generates or can generate, and the analysis above shows which data matter most. Lairage arrival-to-slaughter timestamps (which determine off-feed duration and lairage time) drive roughly 96 per cent of the total figure between them, and are directly measurable. Goad counts, re-stun classifications, HLI inputs and panting scores populate the remaining harms. This is what separates a working tool from a theoretical exercise, with the right timestamps and counts, the model produces a facility-specific modelled estimate capable of structured internal review.

Practically, a facility that collects the right data, implements a change, and re-collects the same data can quantify the welfare impact of that change in the same units. Each is measurable with ongoing data collection, expressed as a reduction in hours of negative affective experience, and reportable. The framework turns AI monitoring from a compliance record into a welfare-improvement instrument, but only to the extent that the enabling data are captured.

9.0 Discussion

The findings of this project converge on a single overarching insight that the value of AI livestock monitoring is determined by the management ecosystem in which it operates, not by the technical capability of the system itself. This section draws together the three analytical strands to articulate the practical implications for the industry.

9.1 Technology Without Governance Is Measurement Without Change

At the facility, the Argus AI system detected events, flagged them for review, and generated trend data across the full operational day. The QA team engaged with it professionally. Yet the evidence from the behavioural data, the survey, and the observed limits of training-based responses to goad use identifies that there are opportunities to further improve data collection and facility responses.

The implication is that AI monitoring implementation must be accompanied by explicit governance design. Who owns the response to detections, what the escalation pathway is from QA finding to livestock team corrective action, how infrastructure issues identified through monitoring are connected to capital planning, and how management visibly demonstrates that acting on monitoring data is expected and valued.

9.2 The Anecdote Problem

A recurring theme across the data review and stakeholder discussions was the prevalence of strongly held operational beliefs that the AI data did not support. The day-of-week goad use pattern was unknown until the data revealed it. The differentiation between precautionary and ineffective re-stuns was not reliably tracked before AI monitoring. The location-specificity of goad use within the race was assumed but not documented.

This is not a criticism of the staff and management involved, it is a structural feature of human observation in complex, high-throughput environments. People work at pace, observe limited windows of the operation, and form patterns from those observations that may or may not reflect what is actually happening across the full-day. AI monitoring replaces these partial observations with complete data. But that advantage is only realised if the operational culture makes room for the data to challenge and update existing beliefs, rather than using the data selectively to confirm them.

9.3 Footage Review as Professional Practice

The examples in Section 6.4 illustrate a principle that deserves specific attention as AI monitoring becomes more widespread in the industry. The natural direction of travel in data intensive systems is toward greater automation and reduced human review. This results in dashboards that replace viewing any footage, summaries that then replace verification, and alerts that automate and so replace investigation. In livestock processing, this direction of travel is risky.

The reasons are structural. The processing environment is complex, variable, and contextually rich in ways that AI systems are not yet capable of interpreting reliably (and may never be). A detection is a signal that a potentially relevant event may have occurred. A human reviewing the footage at that timestamp can determine what actually happened, why, and what response is appropriate. Without this review step, AI monitoring generates data that may be accurate in its classification but misleading in its implications.

This does not mean every detection must be reviewed in real time. It means that the operational design of AI monitoring programmes must include explicit time allocation and role responsibility for footage review, particularly for the event types that are most interpretively ambiguous such as re-stuns reasons, captive bolt use, and location-specific goad events. As AI systems improve their interpretive capability, the scope of required human review may narrow. At this stage of technology development, the industry should treat footage review as an important operational practice, not simply an optional enhancement.

9.4 Infrastructure as a Welfare Lever

The electric goad data in this project makes a case that deserves to be heard at the facilities management and capital planning level, not only at the QA level. Training based interventions produced temporary improvements that did not sustain. The AI data, combined with footage review, points clearly to physical sections of the race where the design constrains animal flow in ways that make goad use more likely.

The welfare improvement available from a targeted race section redesign, potentially involving the position of workers relative to the cattle movements, is likely to be larger and more durable than any number of training sessions. The AI monitoring data makes this case in a documentable, auditable way. Connecting that case to capital planning is a management responsibility.

9.5 The Welfare Footprint Framework as a Strategic Tool

The Welfare Footprint Framework treats intensity and duration as independent axes and keeps them apart. A harm can be brief and intense (small accumulated time, severe per moment) or moderate and sustained (large accumulated time). The rebuilt figures make the distinction stark: roughly 96 per cent of accumulated burden is the moderate-intensity, long-duration lairage experience, while the acute events (goad, handling and the stun) together contribute a fraction of an hour. Because the framework never collapses the intensity categories, both facts remain visible at once.

Existing Audits

Electric goad use and re-stun events receive the most consistent audit attention, yet on these figures they contribute almost nothing to accumulated burden. This is not an argument for auditing them less. A mis-stun matters because it is the only harm capable of Excruciating experience (Section 8.1); its stakes are ethical, not proportional. The correct conclusion is that stun assurance is a non-negotiable baseline that good facilities already meet, while the far larger opportunity for businesses is considering the impact of holding duration and lairage conditions. As this is where the accumulated burden actually sits. In practice, the audit maintains the base while the framework quantifies where the return is.

The current tool and its development trajectory

The intensity distributions are evidence-rated central estimates, and the prevalence and duration inputs are representative values that facility data replace. As AI monitoring populates the off-feed distribution, goad counts, re-stun rates and HLI and panting data, the indicative baseline becomes a measured one and the burden score becomes an operational welfare KPI, refined continuously as data accumulate, exactly as the on-farm and poultry applications of the framework have been.

9.6 Summary of Key Insights

In summary:

- AI monitoring improves visibility, but does not independently drive change;
- Behavioural improvement depends on organisational commitment and system integration;
- Staff adoption is influenced by usability, time, and clarity of responsibility;
- Monitoring systems are most effective when embedded within structured operational processes;
- WFF analysis indicates that, under the illustrative assumptions used here, extended holding and lairage disruption account for roughly 96 per cent of accumulated time in negative affective states. This does not reduce the ethical or assurance importance of acute events; rather, it identifies an additional area for investigation and potential improvement. The framework may help processors and supply chains compare the modelled impact of interventions when the required facility data are available.

10.0 Conclusions

- AI monitoring systems can provide more complete and consistent data than anecdotal observation for understanding animal handling patterns in processing facilities. The full-day, continuous, timestamped nature of AI data reveals patterns that human observation cannot reliably detect.
- The primary determinant of whether AI monitoring drives welfare improvement is management engagement, not technology capability. Without systems that convert monitoring findings into corrective actions for the livestock team, infrastructure decisions, and management accountability, AI monitoring functions as measurement without change.
- Electric goad use at the facility illustrates the infrastructure ceiling on training-based interventions. The AI data supports further investigation of specific race sections and consideration of engineering or operational changes where training alone is insufficient.
- Footage review is an important component of current AI monitoring programmes. Data summaries and detection statistics cannot provide the contextual interpretation

that human review of flagged footage delivers. This is especially important in the complex, variable environment of livestock processing.

- The staff perception survey, while limited in sample size, reveals a meaningful and predictable perception gap in that the operational QA staff experience the burden of false-positive management while management perceives strategic value. Addressing this requires both accuracy improvements and explicit management communication about how findings lead to operational changes.
- The Welfare Footprint Framework provides a potentially useful methodology for estimating the welfare impact of specific interventions. External claims should be limited to modelled estimates unless facility inputs, assumptions, uncertainty and independent review support stronger conclusions.

11.0 Recommendations

For Processing Facilities Deploying AI Monitoring

- Design the management governance structure before deploying AI monitoring, not after. Define who owns the response to detections (not just the review of them), what the escalation pathway is from QA finding to livestock team corrective action, how infrastructure issues identified through monitoring are escalated to capital planning, and what management behaviours signal that acting on monitoring data is expected and valued.
- Establish a formal corrective action register linked to AI monitoring findings. Every validated finding that indicates a systemic issue, rather than a one-off event, should generate a documented response of either a corrective action assigned to a named responsible person with a due date, or a documented decision that no action is required and why. This is the mechanism that converts monitoring from measurement to improvement.
- Allocate protected time for footage review as a designated operational activity. Identify the event types that require footage review (at minimum: re-stuns, captive bolt spikes, location-specific goad events, and any finding that will inform a corrective action) and assign this review to a named role with sufficient time allocation. Do not allow dashboard-only analysis to replace footage review.
- Address the management–staff perception gap through visible action. Hold structured sessions with operational QA and livestock staff that explain how monitoring findings are being used, show specific examples where data led to an operational improvement, and communicate clearly that the system is a welfare and operational improvement tool rather than a surveillance mechanism.
- Prioritise reduction of false-positive detections in collaboration with the AI system provider. Set a target false-positive rate and monitor it as a KPI. False positives are not merely a nuisance, they erode trust in the system among the people closest to it and are the primary driver of the sustainability risk identified in this project.

For the Red Meat Processing Industry

- Develop sector-wide guidance on the governance structures required for effective AI monitoring deployment. Technical guidance on camera placement and system configuration is already available while governance guidance on how to build the management accountability structures that turn monitoring data into welfare improvement is not. This project provides a starting framework.

- Invest in longitudinal welfare burden benchmarking using WFF methodology across multiple sites. WFF welfare burden scores per module provide a basis for industry benchmarking that is currently absent. AMPC is well positioned to lead the development of sector-wide welfare burden benchmarks.
- Treat footage review as a professional standard, not an optional practice. Industry guidance and training programmes for QA staff should explicitly include footage review methodology, event interpretation frameworks, and the competencies required to distinguish infrastructure problems from training problems from equipment issues.
- Ensure future AI monitoring projects build in pre-implementation baselines. This project was constrained by the facility's extended pre-project monitoring history. Future projects should be designed with the monitoring and measurement baseline established before the formal intervention period, enabling cleaner pre/post comparison.
- Explore supplementary sensor integration to improve detection accuracy and reduce false-positives. Smart prodder sensors, acoustic monitors, and break-beam counters can complement CCTV-based AI, improving both attribution accuracy and the confidence of QA staff in acting on detections.

12.0 Project outputs

Output	Description
Retrospective baseline data analysis	Argus AI event data (Aug–Nov 2025) analysed and documented; summary provided to AMPC and the participating facility, December 2025
Staff perception survey	Structured survey developed and administered; n = 4 responses analysed for adoption themes and perception gaps
WFF welfare burden framework	Four-module WFF baseline analysis across receival, lairage, race and stunning/slaughter; published using the illustrative reference dataset in Sections 8.1–8.2
Baseline criteria summary	Retrospective baseline criteria document provided to AMPC and the participating facility; provided as a confidential annex and not included in the public report
Milestone Report (Milestone 1)	Inception and baseline report submitted to and approved by AMPC (December 2025)
Final Report (Milestone 2)	This document
Industry guidance and adoption roadmap	Embedded in Sections 9–11; summarises management governance requirements for effective AI welfare monitoring

13.0 Bibliography

Alonso, W.J. & Schuck-Paim, C. (2021). Pain-Track: a time-series approach for the description and analysis of the burden of pain. BMC Research Notes, 14(1), 229.

Alonso, W.J. & Schuck-Paim, C. (2025). Welfare Footprint Framework: Methodological Foundations and Quantitative Assessment Guidelines. Center for Welfare Metrics.

AMPC (2025). Research Agreement 2026-1031: Behavioural Change Assessment for AI Monitoring. Australian Meat Processor Corporation.

AMPC (2025). Variation Agreement #1: Project 2026-1031. Australian Meat Processor Corporation.

Atkinson, S., Velarde, A., & Algers, B. (2013). Assessment of stun quality at commercial slaughter in cattle shot with captive bolt. Animal Welfare, 22(4), 473–481.

Bond, J., Rumsey, T., & Weinland, B. (1976). Effect of deprivation and reintroduction of feed and water on the feed and water intake behavior of beef cattle. Journal of Animal Science, 43(4), 873–878.

Bourguet, C., Deiss, V., Boissy, A., Andanson, S., & Terlouw, E. M. C. (2011). Effects of feed deprivation on behavioral reactivity and physiological status in Holstein cattle. Journal of Animal Science, 89(10), 3272–3285. https://doi.org/10.2527/jas.2010-3139.

Cockram, M. S. (1991). Resting behaviour of cattle in a slaughterhouse lairage. British Veterinary Journal, 147(2), 109–119.

del Campo Gigena, M., Soares de Lima, J. M., Brito, G., et al. (2021). Effect of finishing diet and lairage time on steers welfare in Uruguay. Animals, 11(5), 1329.

Ewbank, R., Parker, M., & Mason, C. (1992). Reactions of cattle to head-restraint at stunning: a practical dilemma. Animal Welfare, 1(1), 55–63.

Ferguson, D. M., Shaw, F., & Stark, J. (2007). Effect of reduced lairage duration on beef quality. Australian Journal of Experimental Agriculture, 47(7), 770–773.

Gaughan, J., Bonner, S., Loxton, I., Mader, T., Lisle, A., & Lawrence, R. (2010). Effect of shade on body temperature and performance of feedlot steers. Journal of Animal Science, 88(12), 4056–4067.

Gaughan, J., Mader, T., Holt, S., & Lisle, A. (2008). A new heat load index for feedlot cattle. Journal of Animal Science, 86(1), 226–234.

Gaughan, J., Mader, T., Holt, S., Sullivan, M., & Hahn, G. (2010). Assessing the heat tolerance of 17 beef cattle genotypes. International Journal of Biometeorology, 54, 617–627.

Gellatly, D., Meléndez, D. M., Marti, S., Moya, D., Yang, X., Church, J. S., & Schwartzkopf-Genswein, K. S. (2019). Effects of repeated exposure to either energized or non-energized prods on the behavioral responses of beef cattle to handling. Journal of Animal Science, 97(Supplement 3), 7. https://doi.org/10.1093/jas/skz258.011.

George, M., & George, M. (2023). The effect of lairage duration on carcase quality, yield and microbiological status of feedlot cattle (B.FLT.4017). Meat & Livestock Australia.

Gouveia, K., Ferreira, P., Costa, J. R., Vaz-Pires, P., & Costa, P. (2009). Assessment of the efficiency of captive-bolt stunning in cattle. Animal Welfare, 18.

Grandin, T. (1992). Observations of cattle restraint devices for stunning and slaughtering. Animal Welfare, 1(2), 85–90.

Grandin, T. (1997). Assessment of stress during handling and transport. Journal of Animal Science, 75(1), 249–257.

Grandin, T. (1998). The feasibility of using vocalization scoring as an indicator of poor welfare during cattle slaughter. Applied Animal Behaviour Science, 56(2–4), 121–128.

Grandin, T. (2000). Effect of animal welfare audits of slaughter plants by a major fast food company on cattle handling and stunning practices. Journal of the American Veterinary Medical Association, 216(6), 848–851.

Grandin, T. (2002). Return-to-sensibility problems after penetrating captive bolt stunning of cattle in commercial beef slaughter plants. Journal of the American Veterinary Medical Association, 221(9), 1258–1261.

Grandin, T. (2010). Auditing animal welfare at slaughter plants. Meat Science, 86(1), 56–65.

Grandin, T., & Shivley, C. (2015). How farm animals react and perceive stressful situations such as handling, restraint, and transport. Animals, 5(4), 409–421.

Hogan, J., & Phillips, C. (2008). Nutrition and the welfare of ruminants. Annual Review of Biomedical Sciences, 10.

Kiyokawa, Y., Honda, A., Takeuchi, Y., & Mori, Y. (2014). A familiar conspecific is more effective than an unfamiliar conspecific for social buffering of conditioned fear responses. Behavioural Brain Research, 267, 189–193.

Lees, A., Lees, J., Sejian, V., Sullivan, M., & Gaughan, J. (2020). Influence of shade on panting score and behavioural responses of Bos taurus and Bos indicus feedlot cattle to heat load. Animal Production Science, 60(2).

Lefcourt, A., Kahl, S., & Akers, R. (1986). Correlation of indices of stress with intensity of electrical shock for cows. Journal of Dairy Science, 69(3), 833–842.

Mazer, K., Knickerbocker, P., Kutina, K., & Huzzey, J. (2020). Changes in behavior and fecal cortisol metabolites when dairy cattle are regrouped in pairs versus individually after calving. Journal of Dairy Science, 103(5), 4681–4690.

Mitchell, G., Hattingh, J., & Ganhao, M. (1988). Stress in cattle assessed after handling, after transport and after slaughter. The Veterinary Record, 123(8), 201–205.

Munksgaard, L., & Simonsen, H. B. (1996). Behavioral and pituitary adrenal-axis responses of dairy cows to social isolation and deprivation of lying down. Journal of Animal Science, 74(4), 769–778.

Pajor, E., Rushen, J., & de Passillé, A. M. (2000). Aversion learning techniques to evaluate dairy cattle handling practices. Applied Animal Behaviour Science, 69(2), 89–102.

Patterson, D., McGilloway, D., Cushnahan, A., Mayne, C., & Laidlaw, A. (1998). Effect of duration of fasting period on short-term intake rates of lactating dairy cows. Animal Science, 66, 299–305.

Reinemann, D. J., Wiltbank, M. C., Rasmussen, M. D., Sheffield, L. G., & LeMire, S. D. (2003). Comparison of behavioral to physiological response of cows exposed to electric shock. Transactions of the ASAE, 46(2), 507–512. https://doi.org/10.13031/2013.12939.

Schuck-Paim, C., Alonso, W.J., et al. (2026). The welfare impact of heat stress in South American beef cattle and the cost-effectiveness of shade provision. Animals, 16(2), 231.

Sullivan, M., Cawdell-Smith, A. J., Mader, T., & Gaughan, J. (2011). Effect of shade area on performance and welfare of short-fed feedlot cattle. Journal of Animal Science, 89(9), 2911–2925.

Tschoner, T., Kerber, H., Zablotski, Y., Knubben-Schweizer, G., & Feist, M. (2026). Influence of a repeated peripheral electrical stimulation on substance P and cortisol concentrations, and behavior in German Simmental calves. Frontiers in Veterinary Science, 13.

van der Walt, J. G., Hattingh, J., Petty, D., Grobler, M., & Ganhao, M. (1993). Effect of a 72 hour fast on physiological stress indicators in feedlot cattle. Journal of the South African Veterinary Association.

Veissier, I., & Le Neindre, P. (1988). Cortisol responses to physical and pharmacological stimuli in heifers. Reproduction, Nutrition, Development, 28(3A), 553–562.

von Wenzlawowicz, M., von Holleben, K., & Eser, E. (2012). Identifying reasons for stun failures in slaughterhouses for cattle and pigs: a field study. Animal Welfare, 21(S2), 51–60.

Warner, R. D., Ferguson, D. M., Cottrell, J. J., & Knee, B. W. (2007). Acute stress induced by the preslaughter use of electric prodders causes tougher beef meat. Australian Journal of Experimental Agriculture, 47(7), 782–788.

Wertz-Lutz, A., Knight, T., Pritchard, R., et al. (2006). Circulating ghrelin concentrations fluctuate relative to nutritional status and influence feeding behavior in cattle. Journal of Animal Science, 84(12), 3285–3300.

14.0 Acknowledgements

This project was funded by the Australian Meat Processor Corporation (AMPC) through Project 2026-1031: *Behavioural Change Assessment for AI Monitoring of Animal Welfare*. AMPC's support is gratefully acknowledged, along with its ongoing commitment to advancing animal welfare outcomes across the Australian red meat processing industry.

The authors also acknowledge the support of the Commonwealth Government, which contributes to AMPC's research and development activities.

Impetus Animal Welfare Ltd thanks the participating processing facility for providing access to their site, operational data, and personnel for the purposes of this study.

The contribution of SAAS B.V. (Argus) is also acknowledged for provision of the AI monitoring system and technical support associated with its deployment and operation.

The contribution of the Welfare Footprint Institute is also acknowledged for provision of access to their tools and resources to build the evaluation framework and results.

The insights generated in this project reflect the collaboration of all participating organisations and individuals. Any opinions, findings, or conclusions expressed in this report are those of the authors and do not necessarily reflect the views of AMPC or its funding partners.

15.0 Appendices

15.1 Appendix 1 – Data Summary

Not included in the public report. The facility-level event data summary that supported the analysis in Section 6 has been provided to AMPC and to the participating facility as a confidential annex. It is excluded here because it consists of site-specific event counts, device-level and location-level detail, and internal review commentary that are not required in order to interpret or apply the industry findings in this report.

15.2 Appendix 2 – Staff Perception Survey Instrument

Use of Livestock Monitoring AI

This survey aims to understand how livestock AI monitoring systems are being used in practice and identify what would make it easier or more valuable for staff to use.

The goal is not to evaluate individual performance, but to improve the system and how it supports the operation.

Estimated time: 10–12 minutes

Section 1 — Role and Exposure to the System

1. What is your primary role? (Select one)

- QA / Animal Welfare

- Livestock handling

- Production supervisor

- Management

- Maintenance / technical

- Other: ________

2. Which system have you had more contact with? (Select one)

- [Monitoring system A]

- [Monitoring system B]

- Other: ________

3. How often do/did you interact with the monitoring system or its results? (Select one)

- Daily

- Weekly

- Occasionally

- Rarely

- Never

4. Which of the following have you personally done? (Select all that apply)

- Reviewed alerts or detections

- Watched video footage to confirm an event

- Used the data in reporting or KPIs

- Investigated a flagged issue

- Used findings for training or corrective action

- Helped troubleshoot or train the system

- Have not interacted with the system

Section 2 — Perceived Value of the System

5. How useful do you believe the monitoring system could be for the business? (Scale 1–5)

6. What value do you believe the system could provide? (Select up to three)

- Identifying animal welfare risks
- Identifying operational issues
- Internal efficiency (e.g. counting livestock)
- Supporting reporting to customers/regulators
- Supporting staff training
- Improving consistency in procedures
- Reducing disputes about incidents
- I do not see clear value

7. In your experience so far, how accurate are the system detections? (Select one)

- Very inaccurate
- Still training
- Correct enough for use
- Accurate enough for reporting confidently
- Unsure

Section 3 — Practical Barriers to Use

8. How easy is it to find time to review dashboards, alerts, or footage? (Select one)

- Very difficult
- Difficult
- Neutral
- Easy
- Very easy

9. What makes it difficult to use the system? (Select all that apply)

- Not enough time during shifts
- Not seen as a priority by management
- Too many alerts
- Alerts are often false
- Hard to navigate or log into software
- Hard to find relevant footage
- System reliability issues
- Lack of training
- Lack of clear responsibility

Section 4 — Trust and Perception

10. What best describes how you feel about the system? (Select one)

- Useful tool
- Neutral
- Unsure

- Causes unnecessary work
- Concerned about how it may be used

Section 5 — Capability and Training

11. How confident are you in using the system? (Scale 1–5)

12. What training have you received? (Select one)

- Formal training
- Informal explanation
- Learned by trying
- No training

Section 6 — Workflow Fit

13. How well does the system fit into normal work routines? (Scale 1–5)

Section 7 — Organisational Factors

14. How strongly does management support the system? (Scale 1–5)

Section 8 — Improving Adoption

15. What would make the biggest difference? (Select up to three)

- Fewer false detections
- Faster review process
- Clear responsibilities
- Better training
- Stronger management support

Section 9 — Final Reflection

16. If the system disappeared tomorrow, would the operation lose something valuable? (Yes/No)