FREE TEMPLATE

AI Safety Monitoring Pilot Scorecard

Agree the definitions before the trial, set the weights (3 = matters most, 1 = least), then score each system from 1 to 5.

Pilot details

Site and area	
Cameras used (same for every system)	
Dates (from first events)	
Who reviews, and how often	

Event definitions (the same for every system)

Near miss (e.g. person within __ m of a moving vehicle)	
Restricted zones	
PPE required where	
Speed limit, by area	

Scores

What to score (1 = poor, 5 = very good)	Weight	System A	System B	System C	How to check it
Time from camera access to first events					Access date vs first usable events
False events					A person checks a sample
Missed events					Near-miss reports, walks or footage sample
Clip quality: enough to act on					Does the clip show what happened?
Review time					Minutes per week (less is better)
Reports usable without the supplier					Can a manager use them alone?
Privacy settings fit the site					Masking, blurring, retention
Support and changes					Time to add a zone or detection
Events lead to action					Briefings, changes made
Weighted score: (weight × score, added up) ÷ (weights added up)					Leave out rows a system wasn't scored on

Counts

Tally for the trial	System A	System B	System C
Events recorded			
Events checked by a person			
False events in that sample			
Missed events found			
Review minutes per week			
Events that led to an action			

Decision and reasons

Question to finish on: which system's events led to changes on site?