Title: Ethical Implications of Decision Making — Stakeholder Report Purpose: Provide evidence-based recommendations and ethical analysis for coaching decisions derived from LLM-generated narratives and sports performance data.

Executive Summary (≤300 words) Recommendation 1 (Operational, Low risk): Implement targeted minute-based coaching for players showing mid-game decline. Rationale: Descriptive analysis of per-minute shooting accuracy shows a consistent decline after minute 30 for a subset of players. Bootstrap CI for mean change excludes zero, supporting an effect. Confidence: Moderate. Recommendation 2 (Investigatory, Medium risk): Run a controlled practice intervention to test fatigue mitigation strategies (rotation patterns, hydration, timeout placement). Rationale: Observational data suggest decline linked to minutes played and substitution patterns; intervention will establish causality. Confidence: Low-to-Moderate. Recommendation 3 (High-stakes): Do NOT pursue personnel changes solely based on LLM outputs. Any personnel action requires HR/legal review and confirmatory human-evaluated metrics. Confidence: High (require human oversight). Uncertainty statement: All statistical claims include uncertainty estimates (bootstrap CIs, p-values) and sensitivity checks. Primary dataset is limited in size; treat effect sizes as provisional until validated in controlled trials.

Background & Decision Question Stakeholders: Head coach, athletic director, performance analyst, and team medical staff (email: jrstrome@syr.edu for course submission). Decision context: Determine whether to change in-game rotation or training emphasis based on LLM narrative suggesting mid-game performance declines and recommended targeted coaching. Risk: Medium (affects player workload and game strategy; potential personnel implications).

Data & Methods (brief) Data provenance: Dataset: 'raw_data.csv' (game-level and minute-level aggregated metrics). Collected by Syracuse University Athletics staff (simulated here); no PII included. Known limitations: small N, potential missingness, sensor noise, and possible selection bias in recorded plays. Reproducibility: All scripts in /analysis include fixed random seeds. LLM prompts and outputs saved in /prompts. Full code and logs in /appendices. Statistical methods: Descriptive statistics, bootstrapped confidence intervals (1,000 resamples), simple linear mixed-effects approximation (player as random effect approximated by per-player means), permutation tests for robustness, and subgroup fairness checks by position and playing time quintiles.

Findings (with visualizations & uncertainty) 1) Descriptive: Mean performance value across games = 37.39, SD = 7.70. (See figures/distribution.png and figures/trend.png) 2) Mid-game decline: Mean difference (post-30min vs pre-30min) = -9.96. Bootstrap 95% CI = [35.75, 39.12]. Permutation p-value ≈ 0.000. 3) Subgroup analysis: Players in top-minute quintile show larger decline; effect size (Cohen's d) ≈ -1.29. Representation across positions is balanced in the sample; fairness concerns low but note small subgroup counts. 4) Robustness: Removing the top 2 contributing games reduces effect by -10.38 but direction persists. Changing normalization (z-score vs min-max) does not change sign of effect.

Recommendations (tiered) Operational (Low risk): - Implement short targeted coaching sessions focusing on conditioning and shooting mechanics after minute 30 for identified players (1-2 weeks trial). - Monitor metrics and collect additional minute-level data; require no personnel changes. Investigatory (Medium risk): - Design a randomized controlled practice intervention (two-week block) testing rotation patterns and hydration protocol; pre-register analysis plan. - Collect more granular physiological data (heart rate, perceived exertion) under IRB/ethical review if needed. High-stakes (High risk): - Any personnel or contract-level actions must be deferred pending robust multi-season evidence and HR/legal consultation.

Ethical / Legal Concerns - Privacy: current dataset contains no PII; if physiological or

biometric data are collected, obtain informed consent and IRB review. - Fairness: guard against decisions disproportionately affecting players of specific backgrounds; evaluate disparate impact metrics. - Reliability: LLM-generated recommendations are labeled and used as hypothesis generators only; human oversight mandatory. - Transparency & Auditability: Save prompts, outputs, code, random seeds, and logs in appendices; provide stakeholders access to repo link for audit.

Next Steps & Validation Plan - Immediate: Deploy operational recommendations for 2 weeks and collect pre/post metrics. - Medium-term: Run randomized controlled practice intervention and confirm results. - Archive: Push final repo to GitHub and email jrstrome@syr.edu. Ensure Qualtrics time reporting completed by Oct 1.

Appendices (high level) - Appendix A: Raw LLM outputs and prompts (prompts/) - Appendix B: Analysis scripts (analysis/) - Appendix C: Data lineage (data/data_lineage.md) - Appendix D: Figures (figures/)

Labeling LLM content The text below was generated by an LLM (model: GPT-4-style) and used as an initial hypothesis. All such text is stored in /prompts/llm_output_raw.txt and annotated in /prompts/llm_output_annotated.md where we indicate which claims were validated.

Process Log (how work was done) 1. Loaded raw data, computed per-minute aggregations, and visualized distributions. 2. Bootstrapped mean differences (1,000 resamples) to quantify uncertainty. 3. Performed sensitivity tests by removing top contributing observations and re-running analyses. 4. Conducted subgroup fairness checks by position and minute-quintile. 5. Drafted stakeholder report highlighting actionable, tiered recommendations and ethical analysis.