Benchmarking Player Risk Algorithms

JULY 2026

playtech
SOURCE OF SUCCESS

Benchmarking player risk algorithms

What is the topic?

The summarised research is about the use of artificial intelligence (AI) and machine learning (ML) to detect gambling-related risk among customers, and why the gambling sector needs shared benchmarks to evaluate how well these systems work. The research presents a conceptual framework for conducting such benchmarking, in a way that respects data protection and security. Gambling operators, algorithm providers, and interested parties are invited to contact the lead author (kasra.ghaharian@unlv.edu) to explore ways to make this benchmarking a reality.

Why is it important?

Many gambling operators now use, or are expected to use, automated player risk detection systems. These systems analyse player data, such as betting patterns, session length, and product use, to identify customers who may be at risk of gambling harm. In principle, this allows operators to intervene earlier and more precisely.

The challenge is that different operators, vendors, and researchers use different datasets, definitions of risk, modelling methods, and performance metrics. This means that even when a system claims to be "accurate" to a given degree, it is often unclear what that claim means, whether it has been independently tested, and how well it might transfer to different contexts. In gambling harm prevention, the key issue is not only headline accuracy, but whether systems can identify the right people at the right time, while avoiding excessive false positives and false negatives.

This difficulty in performance comparison is particularly detrimental to operators procuring risk detection solutions, but also limits efforts among developers to improve their solutions and efforts by regulators or stakeholders to encourage adoption of high quality tools as part of a broad strategy of supporting players. For instance, the lack of high quality data means regulators don't know what works and are limited in assessing whether their policies are effective or applied effectively.

The absence of benchmarking also exacerbates a wider market problem. Good operators may struggle to prove that their systems are genuinely high quality, while weaker systems may still be marketed as effective. Usage decisions may therefore be driven by price, ease of integration, or sales claims, more than demonstrable performance.

The paper proposes that gambling should learn from other AI fields where benchmarking has helped improve model development. A benchmark, in this context, is a structured way of evaluating models using shared datasets, clearly defined prediction tasks, and agreed performance measures.

Eight prominent academics involved in player risk AI algorithms co-wrote this research, including those from the University of Nevada (Las Vegas), the Division on Addiction at Harvard Medical School, the University of Calgary in Canada, the University of Sydney in Australia, Washington State University, and Focal Research Consultants. Playtech Protect have also been exploring this topic for several years, including launching a public discussion on it via a presentation at the International Association for Gambling Regulators in 2024 (Percy, 2024) and supporting the academic paper that developed the concept (Ghaharian et al., 2026).

What did the research do?

This is a conceptual and policy-oriented paper rather than an empirical study. The authors did not build or test a new risk detection model. Instead, they reviewed the current state of AI and machine learning in player risk detection, identified the barriers to evaluating these systems, and proposed a conceptual framework for benchmarking.

What did the research find?

The paper's primary contribution is sketching how performance benchmarking can work in practice.

The heart of the idea is anonymised player-level data (transaction-level data or session-/day-level aggregated data), split into a training dataset and a testing dataset. A proxy for harm – such as problem gambling screening survey results, clinician judgement, operator exclusion markers, or qualified self-exclusion markers – is shared on the training dataset but withheld on the test dataset. Algorithm providers first prepare their algorithms on the training dataset, before submitting their best estimate of harm for each player in the testing dataset. The benchmarking system then compares those estimates against the withheld harm proxies.

A suite of benchmarks is recommended, rather than a single leaderboard. For instance, different benchmarks may address risk within a session, across a few days, or over several months. Diverse types of player might be included, varying by preferred gambling vertical and jurisdiction, as well as new joiners, infrequent players, established players, and VIPs. Multiple harm proxies should be available, given the limitations inherent in any one proxy. The paper discusses approaches to documentation and data access, including standardised formats, baseline models, defined target variables, and different validation approaches.

A benchmarking system will need to protect privacy, respect commercial sensitivities, and maintain trust. Options include data trusts, regulator-controlled repositories, synthetic datasets, hidden test sets, container-based model evaluation, and tiered validation badges. These mechanisms could allow meaningful evaluation without requiring operators or vendors to disclose all proprietary details publicly. The overall exercise is best led by regulators, independent research institutions, data trusts, or other neutral bodies.

What are the implications for industry and policy?

For industry, the paper suggests a shift from proprietary claims to independently testable evidence. Operators with in-house solutions and solution vendors should expect increasing scrutiny of AI risk detection systems. The sector will need clearer evidence that models work, that they generalise beyond the data on which they were trained, and that they support meaningful harm prevention.

Procurement teams and regulators also have power to drive change. Even short of the full benchmarking vision described, they might prepare a one-off standardised dataset to evaluate competing providers on, perhaps publishing the results as a transparency initiative. They could also request performance data in structured tables to support comparative assessments. Solution providers may also strengthen model governance, with clear documentation, agreed definitions of risk, routine out-of-sample testing, fairness checks, monitoring for data drift, and evaluation of whether model-triggered interventions reduce harm.

Industry inaction may also invite stronger regulatory control. If operators and vendors do not develop credible, transparent, and independently assessable systems, regulators may impose their own models, datasets, or technical standards. Some jurisdictions are already moving in this direction, most prominently in Spain, alongside a broader European trend towards more centralised infrastructure and increased regulatory specificity about models.

References

Ghaharian, K., Dragicevic, S., Percy, C., Nelson, S. E., Murch, W. S., Heirene, R. M., Simeon-Rose, K., & Schrans, T. (2026). The need for benchmarks to advance AI-enabled player risk detection in gambling. Journal of Gambling Studies, 1-26. https://doi.org/10.1007/s10899-026-10483-6

Percy, C. (2024). How can we trust what other people say about AI? International Association of Gambling Regulators (IAGR). Rome, October 2024


playtech
SOURCE OF SUCCESS