

Unfortunately, there are challenges with big data in sport. There are very few shared sports big data for researchers. It happens often that sport institutions or clubs have different sports data, which is mainly in isolation \parencite{bai2021sports} and often licenses for obtaining datasets are needed \parencite{seweryn2023survey}. behalte die struktur aber die sätze sollen noch besser formuliert sein

Gerne! Hier ist deine überarbeitete Passage mit besserem Stil und gleicher Struktur:

Unfortunately, big data in sports faces several challenges. Publicly available sports big data for researchers is still scarce. Frequently, sports institutions or clubs possess disparate datasets that exist largely in isolation \parencite{bai2021sports}, and access to such data often requires costly licenses \parencite{seweryn2023survey}.

Wenn du magst, kann ich die Passage auch in wissenschaftlicherem Stil oder auf Deutsch formulieren.



füge evtl noch einen satz hinzu der sinnvoll sein kann

Gerne! Hier ist die überarbeitete Passage mit einem zusätzlichen, sinnvollen Satz:

Unfortunately, big data in sports faces several challenges. Publicly available sports big data for researchers is still scarce. Frequently, sports institutions or clubs possess disparate datasets that exist largely in isolation \parencite{bai2021sports}, and access to such data often requires costly licenses \parencite{seweryn2023survey}. This lack of openness and standardization hinders the development of comparative studies and the reproducibility of research findings.

Wenn du möchtest, kann ich auch noch einen Vorschlag machen, wie man den Abschnitt stärker mit dem Rest deiner Arbeit verknüpfen kann.



report that this research domain has a lot of room for improvement, where more validation-level and publicly available datasets for replicating research are needed. mache diesen satz evtl. schöner und baue mir eine schöne überleitung zum open source dataset von Pappalardo et al.

Natürlich! Hier ist dein ursprünglicher Gedanke in eine stilistisch flüssige und wissenschaftlich ansprechende Form gebracht, inklusive eleganter Überleitung zum Open-Source-Datensatz von Pappalardo et al.:

This highlights that the research domain still offers substantial room for improvement, particularly with regard to the availability of validation-level and openly accessible datasets that enable the replication and verification of scientific work. In this context, the open-source event dataset provided by Pappalardo et al. \parencite{pappalardo2019public} represents a valuable contribution, as it offers standardized, high-resolution spatio-temporal data across multiple competitions and seasons.

Wenn du magst, kann ich dir auch gleich den nächsten Satz zum Datensatz von Pappalardo formulieren.



formuliere mal den nächsten satz

Gerne, hier ist ein möglicher Folgesatz, der direkt an den vorherigen anschließt:

Its structured format and detailed annotation of match events provide researchers with a solid foundation for developing and evaluating analytical methods in football, fostering reproducibility and comparability across studies.

Wenn du möchtest, kann ich auch eine längere Passage schreiben, z.B. mit Hinweisen auf konkrete Vorteile wie Cross-league analysis, temporal coverage oder Machine Learning-Anwendungen.



highlights that the research domain still offers substantial room for improvement, where more validation-level research is needed and more publicly available datasets to replicate research. In this context \textcite{pappalardo2019public} created the largest open collection of soccer-logs ever released, containing all the spatio-temporal events (passes, shots, fouls, etc.). This publication of this huge dataset could be a milestone, which ... führe diesen satz zuende. was ist sinnvoll

Gerne! Hier ist dein überarbeiteter Absatz, stilistisch verfeinert und mit einem sinnvoll zu Ende geführten Satz:

This highlights that the research domain still offers substantial room for improvement, particularly in terms of validation-level studies and the availability of publicly accessible datasets to replicate research. In this context, \textcite{pappalardo2019public} created the largest open collection of soccer-logs ever released, containing all the spatio-temporal events (passes, shots, fouls, etc.). The publication of this comprehensive dataset could represent a milestone, as it lays the groundwork for reproducible research, facilitates cross-study comparisons, and enables the development of advanced analytical methods in sports science.

Wenn du möchtest, kann ich dir auch einen Übergang zur konkreten Nutzung oder Analyse des Datensatzes in deiner Arbeit formulieren.

