

# High Specificity Of An AI-powered Framework In Cross-Checking Anterior Cruciate Ligament Tear Reports In Public Databases:

Addressing The Need For Improved Data Reliability  
In Media-Based Sports Medicine Research

Pedro Diniz, MD, PhD, Luxembourg LUXEMBOURG

Bernd Grimm, PhD, MEng, Luxembourg LUXEMBOURG

Caroline Mouton, PhD, Luxembourg LUXEMBOURG

Christophe Ley, PhD, Luxembourg LUXEMBOURG

Romain Seil, MD, PhD, Luxembourg LUXEMBOURG



LUXEMBOURG  
INSTITUTE  
OF HEALTH



CHL

Centre Hospitalier  
de Luxembourg

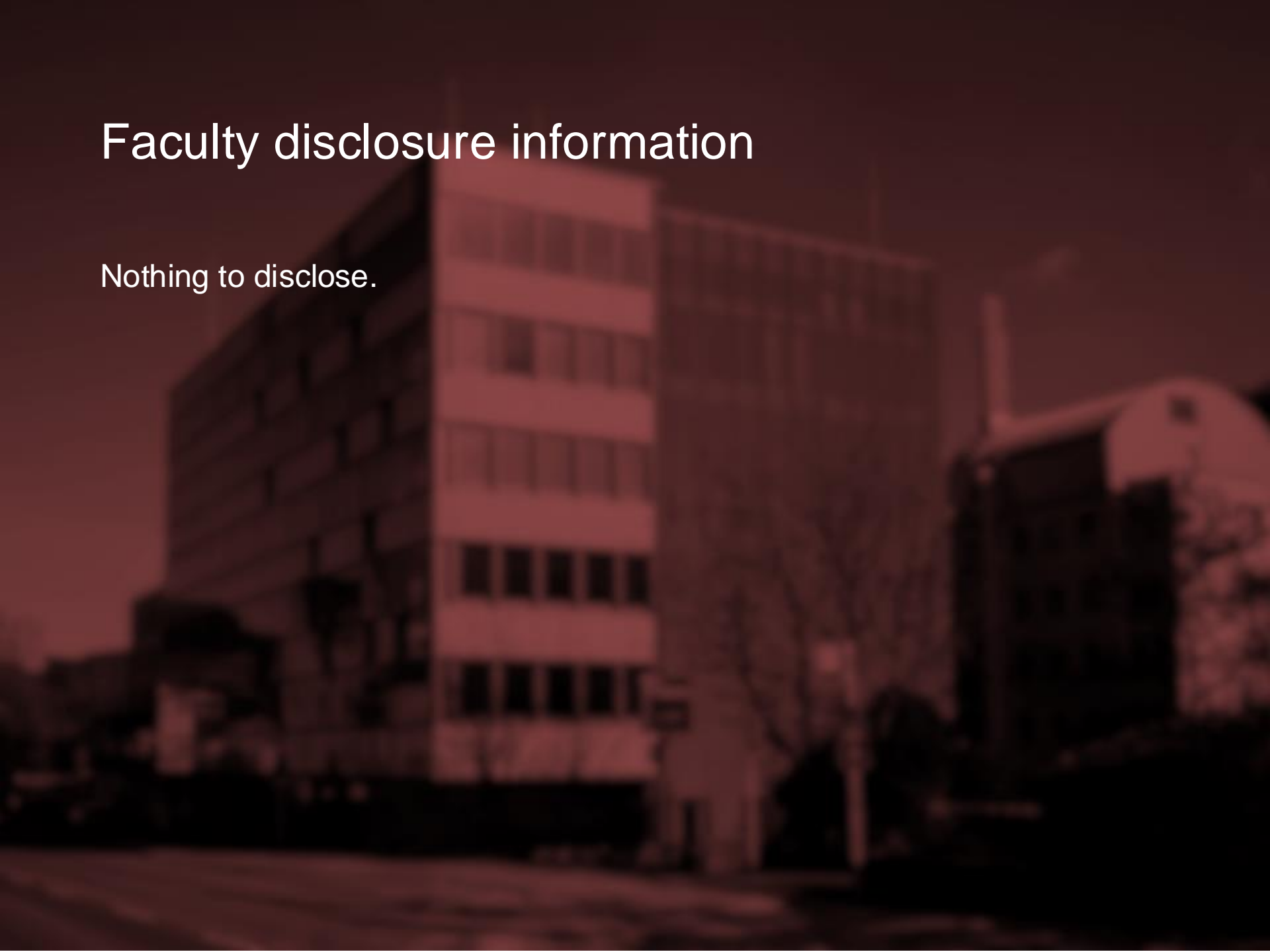


liroms

LUXEMBOURG INSTITUTE OF RESEARCH IN  
ORTHOPEDICS, SPORTS MEDICINE AND SCIENCE

# Faculty disclosure information

Nothing to disclose.



# Introduction

Reliance on public information is increasing in sports medicine research.

(Diniz et al. 2022; Diniz et al. 2023)

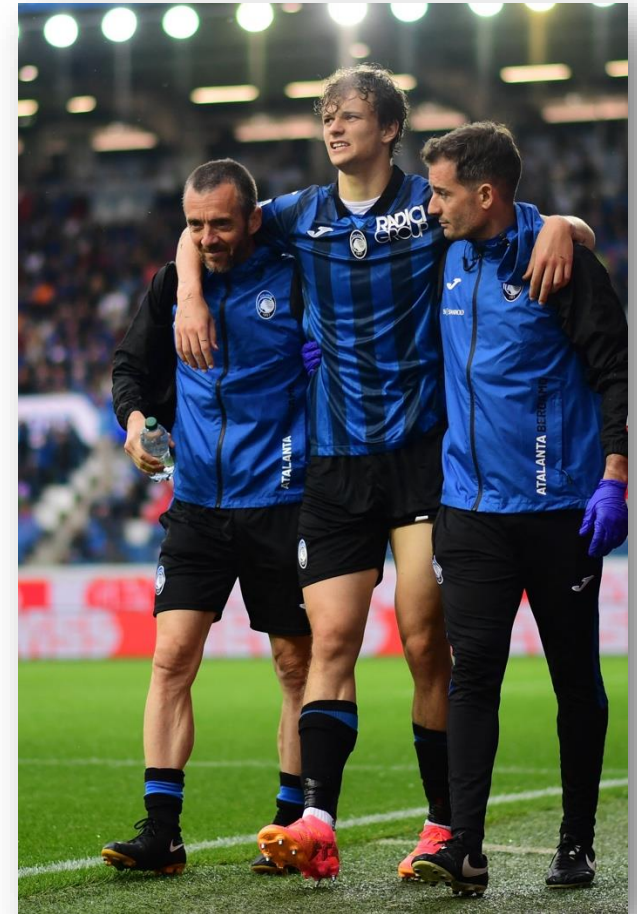
Media-based research enables researchers that do not have direct access to data and opens new possibilities.

Anterior cruciate ligament ruptures (ACL) are a significant concern.

(Della Villa et al 2021)

Serious and career-threatening, considerable time lost, risk for long-term consequences.

(Pinheiro et al. 2023, Kvist and Pettersson 2024)



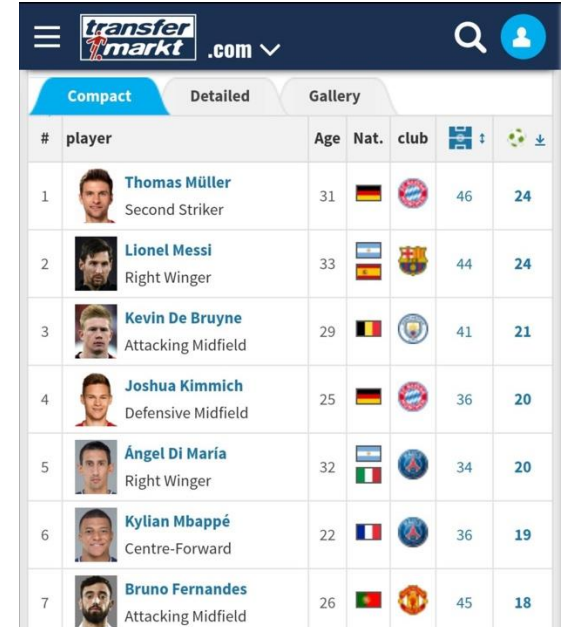
# Transfermarkt

Database includes market values, transfer fees.






















Freely accessible and maintained by the public.

Contains information on injuries, minutes played, match outcomes and much more.

Accurate in 89% of cases for injury location and type.  
(Leventer et al. 2016)



The screenshot shows the Transfermarkt website interface. At the top, there is a navigation bar with the Transfermarkt logo, a search icon, and a user profile icon. Below the navigation bar, there are three tabs: "Compact", "Detailed", and "Gallery". The "Compact" tab is selected. The main content area displays a table of players with the following columns: #, player, Age, Nat., club, and two columns for statistics (46, 24). The table lists seven players: Thomas Müller, Lionel Messi, Kevin De Bruyne, Joshua Kimmich, Ángel Di María, Kylian Mbappé, and Bruno Fernandes. Each player's entry includes a small profile picture, their name, position, age, nationality, club, and the two statistics columns.

#	player	Age	Nat.	club	46	24
1	 <b>Thomas Müller</b> Second Striker	31			46	24
2	 <b>Lionel Messi</b> Right Winger	33			44	24
3	 <b>Kevin De Bruyne</b> Attacking Midfield	29			41	21
4	 <b>Joshua Kimmich</b> Defensive Midfield	25			36	20
5	 <b>Ángel Di María</b> Right Winger	32			34	20
6	 <b>Kylian Mbappé</b> Centre-Forward	22			36	19
7	 <b>Bruno Fernandes</b> Attacking Midfield	26			45	18



# Reliability of public databases

Questions about **validity of public databases**.

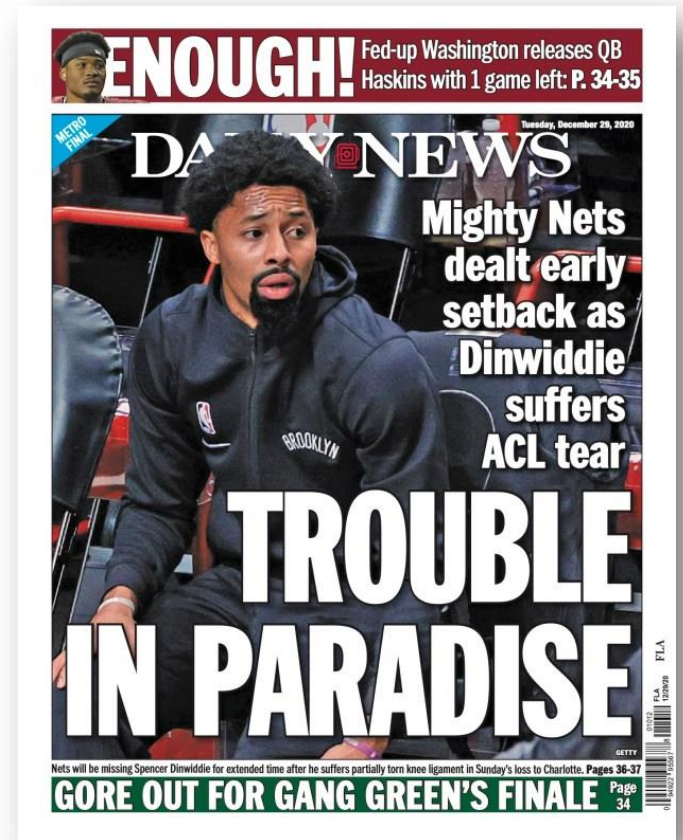
(Carey 2023, Inclan et al. 2022)

**Media reports** on ACL ruptures are **reliable for some types of injuries**.

(Krutsch et al. 2020)

Possible mitigation is to use **Transfermarkt as a starting point** and cross-check injuries elsewhere. However, **manually verifying all ACL injuries** listed in Transfermarkt would be **extremely time-consuming**.

**Hypothesis:** an AI powered framework can cross-check ACL tear-related information from large, publicly available datasets with high specificity and efficiency.



# Materials and Methods

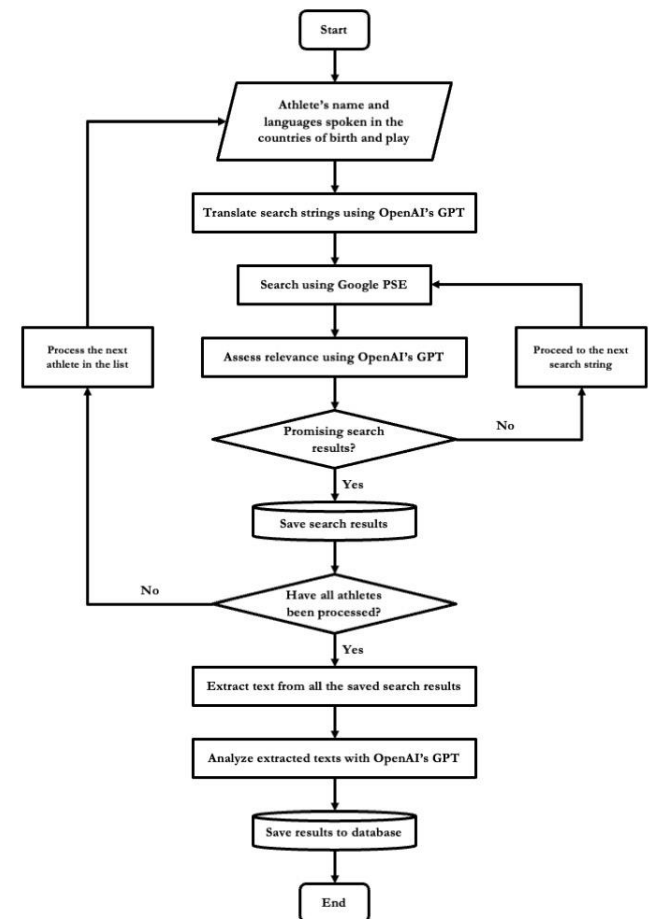
**Database of ACL tears** in male professional footballers from first- and second-tier leagues worldwide (1999-2024) from Transfermarkt.

## AI-powered framework:

Google Programmable Search Engine and OpenAI's GPT-4o-mini.

GPT: translate search queries, appraise search results, and **analyse injury-related information** in search result items (SRIs).

**Specificity** in identifying mentions of ACL tears was the **target performance metric**, and SRIs was the evaluation unit.





Task	Prompt
<b>Translate search query</b>	“You are a sports medicine researcher. Your task is to translate a Google Search query, which aims to find information on soccer injuries, into a specified language. Return only the translated search query.”
<b>Appraise search results</b>	“As a sports medicine researcher, you are examining the results from a Google Search aimed at finding information on soccer injuries, specifically for {player} who may have suffered from an {injury}. Review the provided search links and snippets. Respond with ‘yes’ or ‘no’ to whether any of the content seems to indicate that {player} experienced an {injury}.”
<b>Assess mention of an ACL tear in the player</b>	<p>“As a sports medicine researcher, you need to analyze text that may or may not be specifically about soccer injuries. Your task is to determine if the following text clearly indicates that the player named {player} sustained an injury described as {injury}. Consider the following guidelines:</p> <ul style="list-style-type: none"> <li>- Player Identification: Confirm the player’s full name in the context. Be cautious of players with the same first or last name but different last or first names. Verify the context to ensure the correct player is being referenced.</li> <li>- Injury Confirmation: Look for explicit mentions or strong implications of the injury type. Check for words or phrases that confirm the occurrence of the injury, such as “diagnosed with”, “suffered from”, “sidelined due to”, or similar expressions that unambiguously relate to {injury}.</li> <li>- Contextual Clarity: Ensure that the injury description is not hypothetical or speculative. The context should directly attribute the injury to the player mentioned.</li> <li>- Disambiguation: If multiple players or injuries are mentioned, ensure clarity in the text that directly ties the specific injury to the specified player.</li> </ul> <p>Respond with ‘yes’ if the text clearly confirms that {player} had the injury described as an {injury}. Respond with ‘no’ if the text does not provide clear and direct confirmation.”</p>
...	

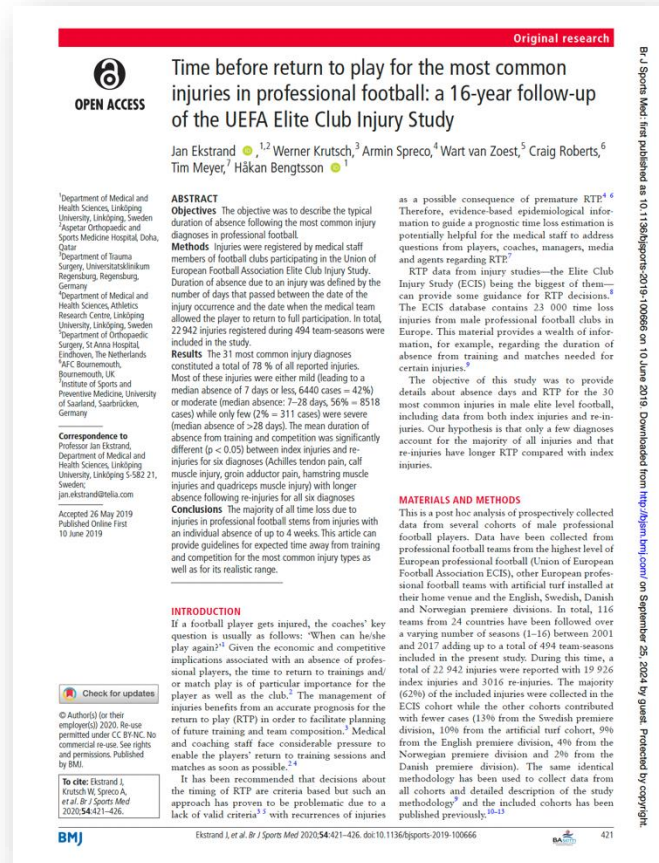
# Materials and Methods

After preliminary testing, 1,522 SRIs were estimated to confirm a 99% specificity for a power of 99.5% and at a 5% significance level.

A researcher extracted injury-related information from each SRI's text, serving as ground truth.

Player age at injury and time until return-to-play were recorded and compared with data from the UEFA Elite Club Injury Study.

GPT processing time and cost were also noted.





# Results

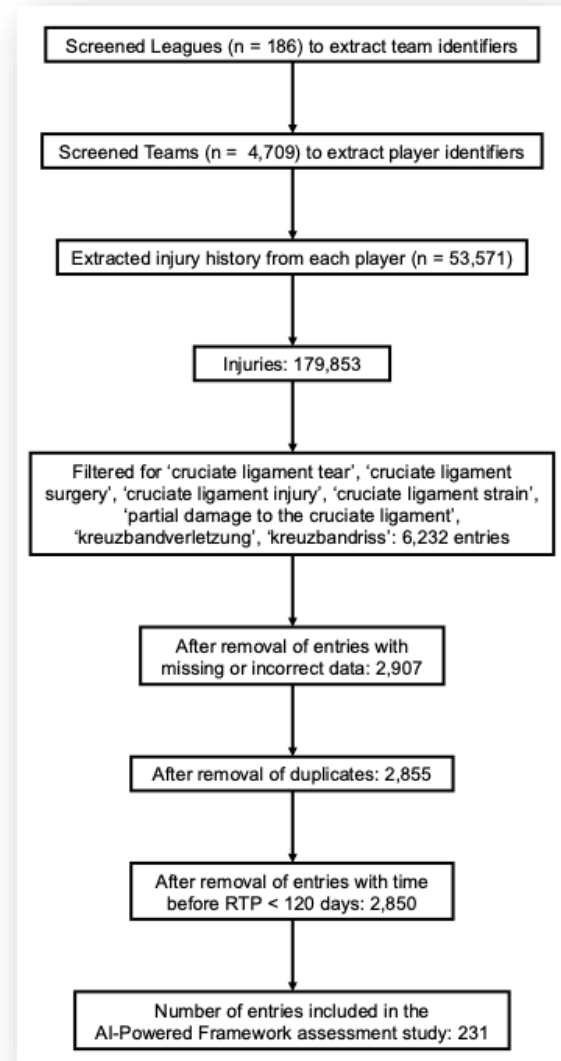
Verification of **231 athletes** yielded 1,546 SRIs.

Human analysis of the SRIs showed that **335 mentioned** an ACL rupture, 6 partial injuries, 51 associated lesions, and in 107 the athlete had or awaited surgery.

*Evidence of ACL rupture in **83 athletes**.*

## Source of SRIs:

- Press articles (n = 853), 55.17%
- Social media (n = 683), 44.18%
- Other (n = 6), 0.39%.
- Official announcements (n = 3), 0.19%
- Interviews (n = 1), 0.06%



# Results

Task	Performance metrics (%)			
	Specificity	Sensitivity	Accuracy	Precision
Assess mention of an ACL tear in the player	99.3	88.4	96.9	97.0
Assess whether the player had a partial or complete injury	99.9	66.7	99.8	80.0
Assess whether the player had associated knee injuries	98.9	84.3	98.4	71.7
Assess whether the player had or was scheduled for surgery	98.4	76.6	96.9	78.1

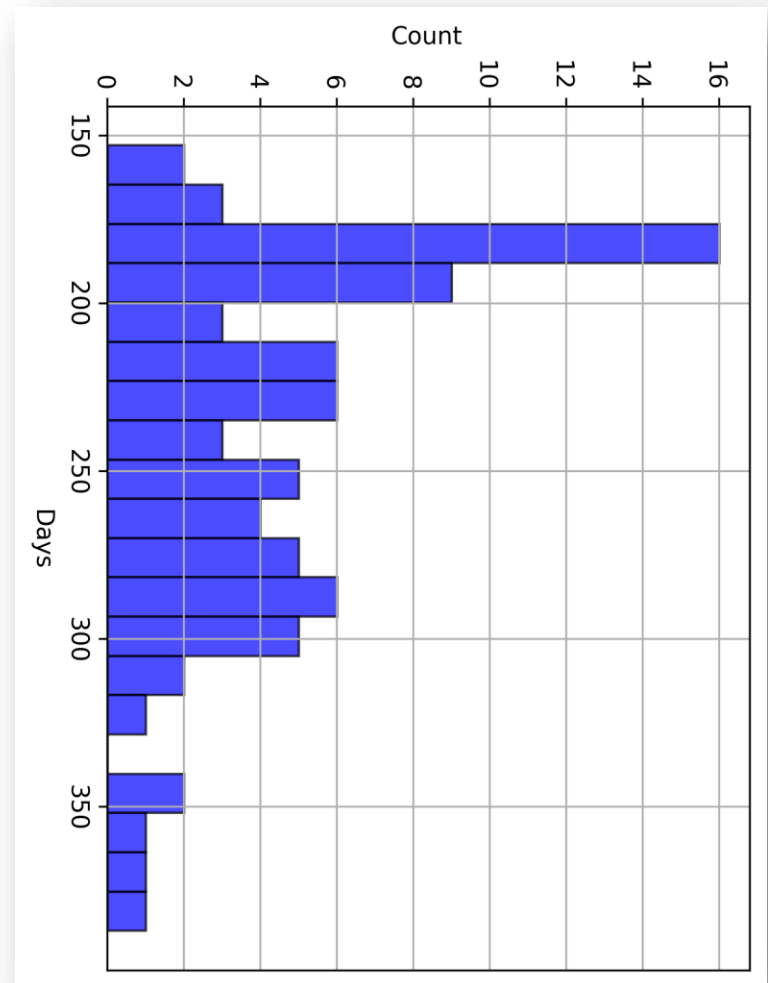
# Results

The mean **age at rupture**:  
26.6 years (SD: 4.6, 95% CI: 25.6-27.6).

Median **return-to-play time**:  
225 days (IQR: 96, 95% CI: 209-251).

**Comparable to reports** using data  
from the UEFA Elite Club Injury Study.

GPT processing took 41 minutes  
and cost around 5 USD.



# Discussion

AI-powered framework: **high specificity** cross-checking male professional football ACL tear reports from a public database.

May significantly **reduce manual workload and enhance reliability** of sports medicine research using public datasets.

Ages and RTP times aligned with UEFA Elite Club Injury Study, **supporting data credibility**: highlights the potential for research when clinical datasets are limited.

Limitations: lower accuracy for partial tears and associated injuries. Unknown generalisability to female athletes, other sports, or injuries.

Could **fine-tuning** the language model **improve performance**?

# Conclusion

Media-based research is a relevant opportunity but reliability must be ensured

AI-powered framework: high specificity cross-checking male professional football ACL tear reports from a public database.

Future research: explore application in other injuries and sports; expand the framework's capabilities.

# References

Diniz P, Abreu M, Lacerda D, Martins A, Pereira H, Ferreira FC, Kerkhoffs GM, Fred A (2022) Pre-injury performance is most important for predicting the level of match participation after Achilles tendon ruptures in elite soccer players: a study using a machine learning classifier. *Knee Surg Sports Traumatol Arthrosc* 30(12):4225–4237

Diniz P, Lacerda D, Mendes B, Pereira H, Ferreira FC, Kerkhoffs GMMJ (2023) Return-to-performance in elite soccer players after Achilles tendon ruptures: a study using a weighted plus/minus metric and matched-control analysis. *Knee Surg Sports Traumatol Arthrosc* 31(12):6059–6068

Della Villa F, Häggglund M, Della Villa S, Ekstrand J, Waldén M (2021) High rate of second ACL injury following ACL reconstruction in male professional footballers: an updated longitudinal analysis from 118 players in the UEFA Elite Club Injury Study. *Br J Sports Med* 55(23):1350–1357

Pinheiro VH, Borque KA, Laughlin MS, Jones M, Balendra G, Kent MR, Aigaoankar R, Williams A (2023) Determinants of Performance in Professional Soccer Players at 2 and 5 Years After ACL Reconstruction. *Am J Sports Med* 51(14):3649–3657

Kvist J, Pettersson M (2024) Knee-Related Quality of Life Compared Between 20 and 35 Years After an Anterior Cruciate Ligament Injury Treated Surgically With Primary Repair or Reconstruction, or Nonsurgically. *Am J Sports Med* 52(2):311–319

Leventer L, Eek F, Hofstetter S, Lames M (2016) Injury Patterns among Elite Football Players: A Media-based Analysis over 6 Seasons with Emphasis on Playing Position. *Int J Sports Med* 37(11):898–908

Carey JL (2023) The Critical Reader: Data Sources. *Am J Sports Med* 51(12):3103–3105

Inclan PM, Chang PS, Mack CD, Solomon GS, Brophy RH, Hinton RY, Spindler KP, Sils AK, Matava MJ (2022) Validity of Research Based on Public Data in Sports Medicine: A Quantitative Assessment of Anterior Cruciate Ligament Injuries in the National Football League. *Am J Sports Med* 50(6):1717–1726

Krutsch V, Grechenig S, Loose O, Achenbach L, Zellner J, Striegel H, Alt V, Weber J, Braun M, Gerling S, Krutsch W (2020) Injury Analysis in Professional Soccer by Means of Media Reports – Only Severe Injury Types Show High Validity. *Open Access J Sports Med* 11:123–131

Ekstrand J, Krutsch W, Spreco A, Zoest W van, Roberts C, Meyer T, Bengtsson H (2020) Time before return to play for the most common injuries in professional football: a 16-year follow-up of the UEFA Elite Club Injury Study. *Br J Sports Med* 54(7):421–426