

The logo for the 2024 Embedded VISION Summit is centered on the left side of the slide. It features a white octagonal background with a colorful, multi-layered border in shades of purple, blue, green, yellow, and orange. The text "2024" is at the top, "embedded" is below it, "VISION" is in large, bold, dark blue letters with a gradient, and "SUMMIT" is at the bottom in a smaller, dark blue font.

2024
embedded
VISION
SUMMIT®

Using Vision Systems, Generative Models and Reinforcement Learning for Sports Analytics

Mehrsan Javan

Chief Technology Officer

Sportlogiq

SPORT
LOGiQ

Sport Analytics Is About Insights Extracted from Data

- Data Generation: Cost efficient, scalable, and robust data acquisition from videos using fully automated/AI-assisted systems
- Insights Generation: Descriptive and predictive analytics and relevant insights for users at scale for performance evaluation, scouting, media content creation, prediction, etc.



Sport Analytics in Different Leagues

- Top tier pro leagues (NHL, NFL, EPL)
 - Accurate raw datapoints from different sources
 - Advanced analytical skills and deep sports knowledge, inhouse data science teams
- 2nd tier pro & draft eligible leagues
 - Partial and almost accurate raw datapoints
 - Domain expertise with limited analytical knowledge
- 3rd tier leagues and youth games
 - Partial, incomplete and inaccurate raw data
 - Limited domain expertise and almost no analytical knowledge

Analytics in Professional Leagues

- Analytics foundations are based on uniform and accurate input data
- Catered for scouts, players, coaches, GMs, and media
 - Player/team metrics and physical data
 - Player archetypes, search and recommendations, scouting tools
 - Expected value models and reinforcement learning tools for strategy evaluation

Sample NHL Player Card in Sportlogiq app



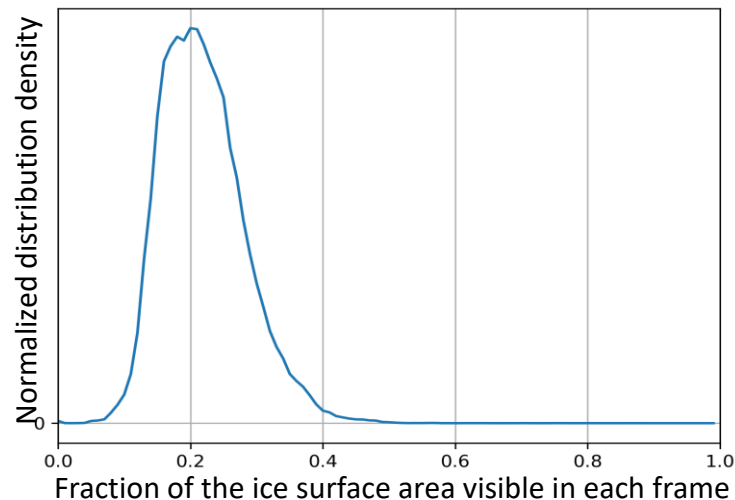
Extending Analytics from Pro Leagues to Youth Sports Is Not Straightforward

- Data incompleteness — regardless of how strong our computer vision systems are, we are limited to what the camera sees
 - Any useful analytical insight requires reliable and uniform input data
 - Pro level analytics models don't work out of the box with youth data
- Insights relevance — what is relevant in the pro leagues may not be relevant for the youth sports users
 - Users neither have the sports knowledge nor analytics backgrounds
 - Educating users takes a long time, in the meantime they need a valuable product

Data Incompleteness and Inaccuracy — Partial Observations

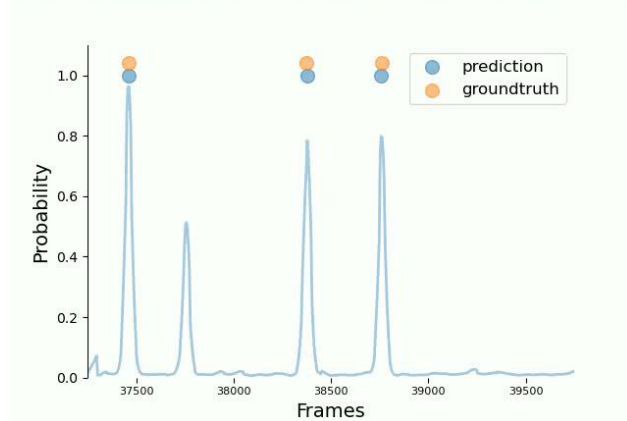
- Partial player observations
 - Player location data exists for a subset of players that are visible in the video
 - Not necessarily all players are observed and tracked for the whole games

Distribution of the fraction of the ice area covered in the camera field of view

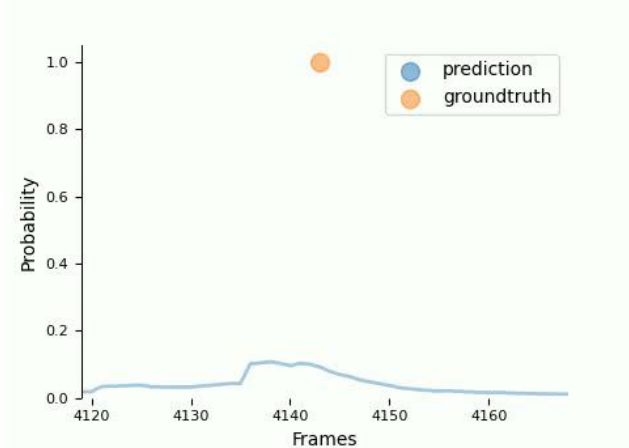
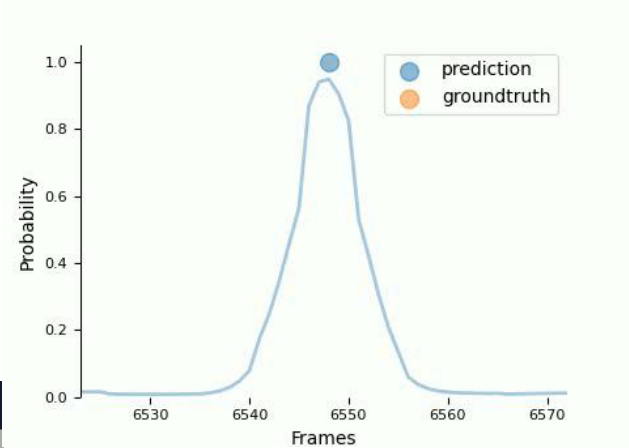


Data Incompleteness and Inaccuracy — Partial Observations

- Partial game events and play sequence data
 - Players actions and game events have a higher false detection rates compared to pro leagues
 - Some game events are not collected in youth games

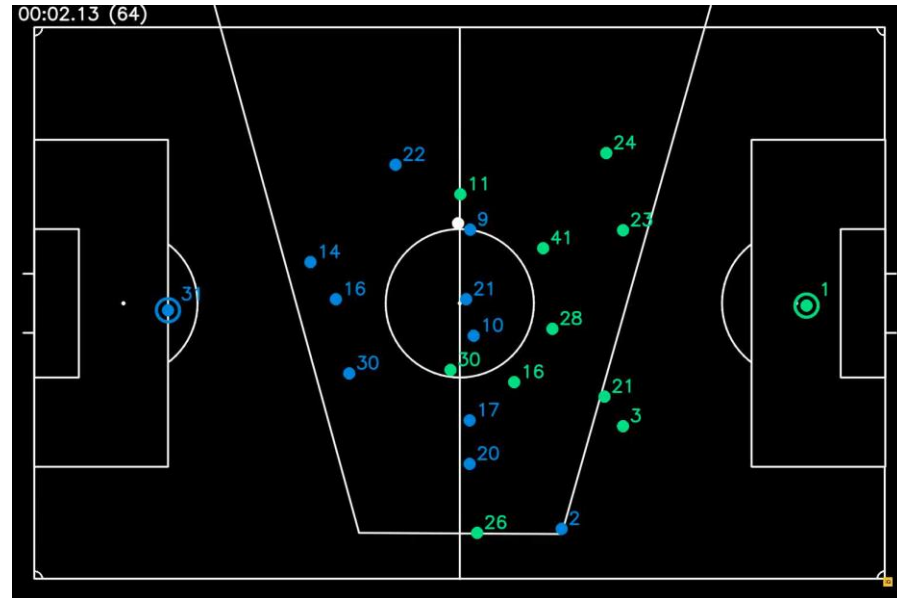


Data Incompleteness and Inaccuracy — Partial Observations



Generating Complete Data from Partial Observations

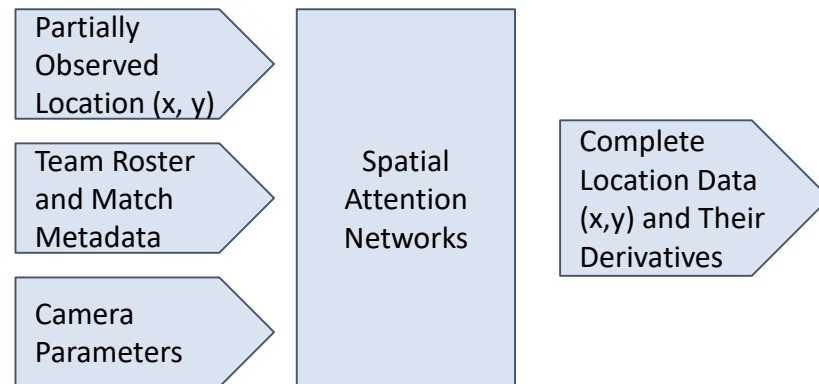
- Use of Generative AI for data creation
 - Complete player locations for physical data and metrics generation
 - Plausible sequences of on puck/ball actions by filling the missing actions and correcting mistakes



Generative AI for Data Creation — Trajectory Data

- Input tracking data is a partial set of observed location data extracted using computer vision tracking techniques
- Generated output is a complete set of player location data
 - Auxiliary outputs for instantaneous velocity and acceleration
 - Player physical metrics
- Group behavior modeling to capture team tactics remains a challenge

Tracking Data Generation



- Problem Statement: Given a set of partially accurate and incomplete sequences of on puck/ball actions, generate a plausible play sequence for the whole game
- A play sequence is a complete set of events, with the event type, other attributes, outcome, time, location on the ice, and global game state
 - An abstract example of a play sequence without attributes: {Faceoff, Loose Puck Recover, Pass, Reception, Shot, Blocked Shot, Loose Puck Recovery, Shot, Goal, Whistle}

- Modelling approach is inspired by language models
 - Each event can be equivalent to a sentence in a document
 - A document is an episode of play in hockey game starting with a faceoff and ending with a whistle

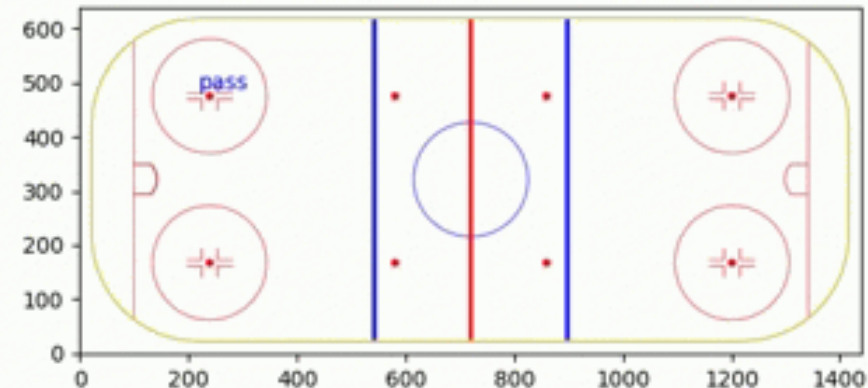
Example Event Token

```
{  
  "tag": "playerEvent", "period": 2,  
  "periodTime": 721,  
  "teamId": 15, "playerId": 1330,  
  "xAdjCoord": 37.92, "yAdjCoord": 39.48,  
  "playersOnIce":  
  [384,380,1321,1330,1753,11581,627,305,1348,1964  
  ,1968,1977],  
  "teamInPossession": 15, "scoreDifferential": -  
  1,  
  "playZone": "oz", "playSection": "eastPoint",  
  "eventName": "pass", "outcome": "successful"}  
}
```

Generative AI for Data Creation — Game Events

- Base model: Generative Pretrained Transformer (GPT) which captures the causality in the play sequence data
 - Over 10 seasons of data from 5 professional hockey leagues were used for training the base model
- The model was then tuned with limited number of youth games (a few hundred)
- GPT model passed both automated and the human QA systems

Game Event Sequence Creation

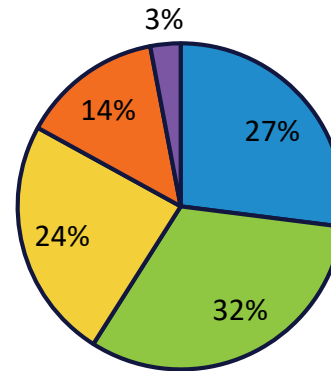


We got the data, but what is the end product?

Relevant Insights for the Youth Markets

- Userbase Segmentation
 - 10% team and personal coaches
 - 57% family members who are not parents and are over 56 or mention their grandchild
 - 33% parents and players
- Users are not analytically oriented
 - 27% have no understanding of stats
 - 32% understand basic stats
- 48% not interested in learning analytics

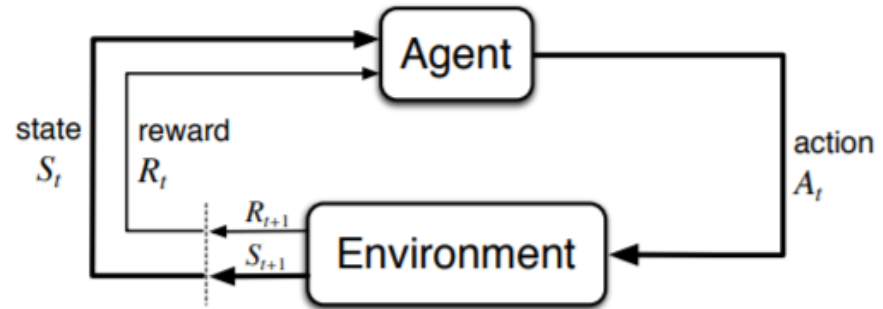
Users' Sport Analytics Knowledge



- I don't understand analytics at all
- I can follow basic stats and analytics discussions
- I can understand visualizations and chart analytics
- I can easily analyze and explain chart analytics and data visualizations
- I examine raw data to create detailed analyses and strategic insights

What Is Behind Pro Leagues Analytics?

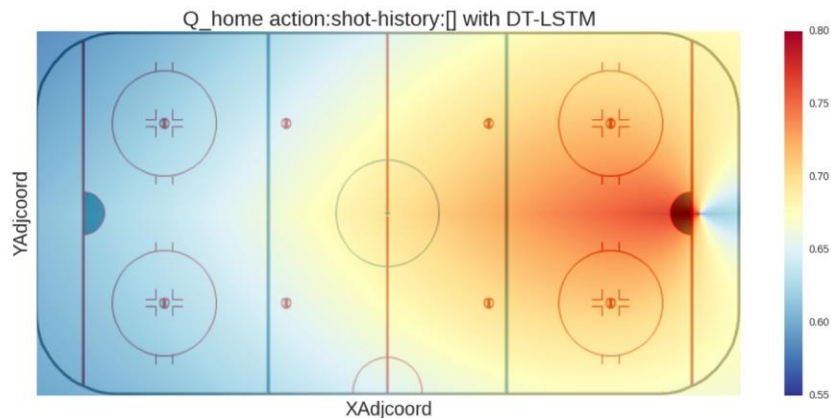
- The most powerful and sophisticated tool in sport analytics is Inverse Multi-Agent Reinforcement Learning (MARL)
- A game is a sequence of <state, action> with a reward signal, each game state has a quantitative value
- All advanced player metrics and profiling are based on the MARL systems
- Inverse MARL can tell us what a team/player is trying to optimize



The Use of Pro Leagues MARL for Youth Games

- A tool already exists that can assign quantitative values to each game state
- Reward signal can get modified to measure changes in game states for each team based on offense/defense objectives
- Significant changes in the game state values are important moments in the game – highlight-worthy segments
- Game/player highlights can get generated automatically from videos

Example: value of shots at different locations on ice given a specific game context



User Specific Highlights

- Instead of a generic game highlight, a user specific highlight can get created for each player
- Using a variety of reward signals, based on a player position (offense, defense) a highlight reel of offensive or defensive plays can get created
- Highlights are accompanied with light stats
- Highlighted segments are marked based on how interesting they are

Highlight Example

ER 11, 2023 - 7:00PM EST
Dean-Mance Arena

Hawks

#77 - M. Wasserman

Game top speed: 6.5mph
#77 M. WASSERMAN

Highlight ID	Event Type	Rating	Time	Period	Date
2	RUSH	☆☆☆	0:43	1st period	2023-11-11
3	SHOT	☆☆☆	1:04	1st period	2023-11-11
4	GOAL	☆☆☆	1:05	1st period	2023-11-11
5	SHOT	☆☆☆	1:04	1st period	2023-11-11
6	GOAL	☆☆☆	1:05	1st period	2023-11-11
7	SHOT	☆☆☆	1:04	1st period	2023-11-11
8	RUSH	☆☆☆	0:43	1st period	2023-11-11
9	SHOT	☆☆☆	1:04	1st period	2023-11-11
10	GOAL	☆☆☆	1:05	1st period	2023-11-11

13 clips
2:53

Share Download

Youth Sports Product Development — Conclusions

- End users in youth sports want simple yet valuable insights
- Pro level insights won't be useful in the youth sports in their current shape and form
- Producing reliable sport analytics content requires vast amount of uniform and accurate data
 - Generative AI can fill the gaps in the input data
 - Complete player location data can get generated from partial observations
 - A plausible play sequence for on puck action can get generated from inaccurate and partial play sequence data
 - MARL can be used to generate player specific highlights
- Sport data is multimodal and requires multi-modal generative AI tools

For More Information

Videos

All videos and demos are available at
<https://sportlogiq.com>

References

Liu et al., Learning agent representation
for ice hockey, NeurIPS2020

M. Horton, Learning feature representation
from football tracking, Sloan sport analytics
2021

US Patents and pending applications
11,130,040; 18/529,204; 17/445,354;
17/817,454