Project Proposal: AI-Powered Real-Time NFL Strategy Report

Team Information

Team Name: Data Depth Explorers

Team Members:

- Krutarth Lad - LLM Integration, Machine Learning & RAG Implementation
- Jainil Patel - Graph Neural Network(GNN), NLP, Vector Database
- Jason Jordan - LLM Integration, Computer Vision, Domain knowledge
- Harshita Behera - YOLO, Data visualization

Individual Responsibilities:

- Krutarth Lad: Implement Retrieval-Augmented Generation (RAG), real-time sentiment analysis, and AI-driven strategy reports.
- Jainil Patel : Implement Data Pipeline setup, biomechanical modeling
- Jason Jordan: Data pre-processing and integration, image/video analysis
- Harshita Behera: YOLO-based player tracking, data visualization

Skillsets:

- Machine Learning (Regression, Classification, Neural Networks, Graph Neural Networks)
- NLP (Transformers, Sentiment Analysis, RAG)
- Computer Vision (YOLO, MediaPipe, Pose Estimation)
- Data Engineering (AWS, Vector Databases, nflverse, Pro Football Reference API)

Significance & Motivation

Project Importance

- NFL teams and coaching staff are increasingly turning to data-driven insights to enhance performance and gain a competitive edge. However, existing data systems often provide fragmented insights that don't translate effectively into actionable strategies. This project integrates multi-modal data sources—including player tracking, video analysis, weather conditions, news, and social media—with advanced AI techniques to generate real-time, adaptive strategy reports.

- Coaches require dynamic, live reports that factor in player mobility, biomechanics, weather, and psychological factors to make informed, data-driven decisions during games. While current tools mainly focus on post-game analysis, our solution delivers real-time, in-game updates, improving decision-making on the field.

Problem Addressed

- Current strategy and play prediction tools are designed for post-game analysis, providing limited real-time insights that could influence decision-making during active gameplay. Additionally, integrating and interpreting the vast variety of data sources—including tracking data, video, social media, and weather reports—presents significant challenges.
- Our system leverages generative AI to produce dynamic, real-time reports, allowing coaches and analysts to adapt strategies instantly. By addressing the need for real-time strategic adaptation, our solution enables teams to process complex data streams efficiently and make live game adjustments, moving beyond reliance on post-game evaluations.

Applicable

- **NFL Teams & Coaching Staff**: Real-time, AI-driven insights to enhance decision-making and improve game strategies.
- **Players**: Identification of individual performance trends, fatigue risks, and matchup vulnerabilities.
- **Sports Analysts & Scouts**: Deeper insights into team tendencies, play-calling strategies, and opponent weaknesses.
- **Sports Betting & Fantasy Football Platforms**: More accurate predictive analytics for improved decision-making in betting odds and fantasy team selections.

Objective & Innovation

Main Goal:

The main goal of this project is to develop an AI-powered NFL scouting report generator that provides real-time, data-driven insights for teams. This system will integrate player tracking, opponent analysis, social media sentiment, weather conditions, and injury risks to generate adaptive strategy recommendations before and during games. By leveraging AI, teams can optimize play formations, defensive schemes, and player utilization to maximize performance and minimize risks.

Approach Novelty:

Unlike traditional post-game analysis, our system will dynamically update and adapt throughout the game. By combining real-time tracking data, social media sentiment, video-based player detection (YOLO), weather forecasts, and news, the system will provide insights and predictions with unmatched relevance and immediacy. We'll employ a Retrieval-Augmented Generation (RAG) approach to constantly feed new, relevant data into the AI model.

Innovations and Techniques We Will Implement:

- **AI-Driven Play Prediction**:
  - Using LLMs (DeepSeek-7B) fine-tuned on NFL playbooks to anticipate opponent play calls based on real-time formations.
- **Player Detection**:
  - YOLO for real-time video analysis to locate players and the ball, enhancing play prediction.
  - MediaPip for detecting player pose and orientation which will provide tactical insights for in-game adjustments.
- **Multimodal Data Fusion:**
  - We combine structured data (player tracking, weather, injuries) with unstructured data (news, social media sentiment, fan/coaching perspectives). This comprehensive approach ensures a more holistic analysis.
- **RAG for Dynamic Reports:**
  - Instead of static scouting reports, we use Retrieval-Augmented Generation (RAG) to fetch the most relevant real-time information and generate customized strategy reports for each game scenario.
- **Sentiment-Aware Strategy Adjustments**:
  - Using NLP and clustering algorithms to extract player morale, coaching sentiment, and fan expectations from social media.
- **Environmental Factor Integration** :
  - Analyzing weather conditions and injury risk models to predict performance impacts.
- **Dynamic AI-Generated Reports** :
  - Providing adaptive, human-readable reports that update in real-time, offering actionable recommendations to coaches.

Role of Generative AI

Generative AI Integration:

- Generative AI will be used to create the final strategy report. It will process inputs such as historical data, real-time tracking, player location, social media sentiment, and news updates to generate actionable insights for coaches, analysts, and even

viewers. The system will leverage large language models (LLMs) to generate natural language text that explains optimal strategies and adjustments.

Models, Architectures, and Techniques:

- **Retrieval-Augmented Generation (RAG) for Strategy Reports:**
  - Continuously fetch news, social media, injury reports, and filter relevant updates.
  - Use DeepSeek-7B (fine-tuned on NFL playbooks) to generate natural language strategy insights.
- **Pose-Based Analysis:**
  - Extracts real-time biomechanics (hip angles, stance width, reaction time) to predict player matchup advantages.
- **Multi-Modal Fusion:**
  - Combine tracking data + text analysis + real-time pose analysis using a Graph Neural Network (GNN).
- **Real-Time In-Game Updates:**
  - AI dynamically adjusts offensive/defensive recommendations based on player fatigue, weather changes, and opponent adaptation.

Generative AI Applicability:

Generative AI automates the strategy formulation process by analyzing vast amounts of data and providing real-time recommendations. It eliminates manual review and speeds up decision-making, enhancing the coaching staff's ability to make adjustments mid-game. By leveraging temporal-based models for mobility predictions, the AI can better anticipate player movements, allowing for more accurate predictions and optimal strategies.

Related Work (Literature Review)

Existing Research and Gaps

1. **Big Data Bowl 2024 - NFL Player Tracking**
   - **Findings:** Used Next Gen Stats player-tracking data to predict play outcomes.
   - **Methodology:** Machine learning models, including XGBoost and LSTMs, to classify play types.
   - **Gap:** Focused on post-play analysis rather than real-time decision-making and lacked integration of external factors like weather and sentiment analysis.
   - **Citation:** Big Data Bowl
2. **AI assistant for (European) football tactics**

- **Findings**: Able to predict outcome of corner kicks and generate tactical variations that were favoured over existing expert tactics 90% of the time.
- **Methodology**: GNN to analyse player spatial relationships and predict outcomes and Autoencoder to simulate new positionings to increase the probability of success.
- **Gap**: Only focuses on european football as well as only one game-time scenario (corner kick)
- **Citation**: Wang, Z., P. Velickovic, D. Hennes, et al., (2024) TacticAI: an AI assistant for football tactics. *Nature Communications*. 15. 1906

3. **QB-GPT: Generative AI for NFL Play Prediction**
    - **Findings**: GPT-2 was fine-tuned to generate potential NFL plays based on previous play sequences.
    - **Methodology**: Used transformer-based models to synthesize plays from structured play-by-play data.
    - **Gap**: Lacked multi-modal integration of visual tracking, weather impact, and social media insights.
    - **Citation**: Towards Data Science
4. **Social Media Driven Predictions of Athlete Performance**
    - **Findings**: Assessed NBA player performance based on features extracted from pre-performance social media posts from the distinct player and about the player. The number of Twitter postings received by NBA players contributed significantly to performance predictions.
    - **Methodology**: Social media postings were obtained from the R package *academictwitteR* and pre-processed into features before training on a random forest model.
    - **Gap**: Outcome did not consider relevant context of postings and assigned words/phrases with positive/negative scores without taking into account the possibility of dual meanings. Sentiment analysis through enhanced NLP embeddings would likely improve these results.
    - **Citation**: Dreyer, F., J. Greif, K. Gunther, et al., (2022) Data-Driven Prediction of Athletes' Performance based on their Social Media Presence. *Discovery Science: 25th International Conference, DS 2022, Montpellier, France, October 10–12, 2022, Proceedings*. pp. 197-211
5. **Pose Estimation for Sports Analytics**
    - **Findings**: Pose estimation techniques like OpenPose and MediaPipe have been used for biomechanics analysis in basketball and soccer.
    - **Methodology**: Keypoint detection models analyze player stances and movements.
    - **Gap**: Limited application in American football, especially for offensive/defensive line matchups.

- **Citation**: Tuyls, K., S. Omidshafiei, P. Muller, et al., (2021) Game Plan: What AI can do for Football, and What Football can do for AI. *Journal of Artificial Intelligence Research*. 71. 41-88.

Challenges:

- **Real-Time Data Integration**: Integrating various data streams (tracking, social media, weather, news) in real-time presents a significant challenge.
- **Data Quality and Noise:** Ensuring that the data, especially from social media and news sources, is clean, relevant, and meaningful despite potential noise or irrelevance.
- **Model Adaptation:** Continuously updating AI recommendations based on the dynamic nature of the game and evolving contexts.
- **Player Mobility Prediction Accuracy:** Using YOLO, or other temporal models to predict player mobility while ensuring the model generalizes effectively across diverse game situations without overfitting.

Solutions:

- Implement robust data preprocessing pipelines to clean and validate data, ensuring only relevant and high-quality information is used.
- Utilize advanced NLP techniques (e.g., RAG) to keep the system updated with the latest, relevant insights from news and social media.
- Work closely with football strategy analysts to validate AI-generated insights and adjust models as needed for real-time strategy application.
- Regularly retrain temporal models using new player mobility data to enhance accuracy and ensure the models generalize well across different game scenarios.

Potential Dataset

- **Next Gen Stats (NFL Kaggle)**: Contains player mobility, play outcomes, and time-based tracking data.
- **Pro Football Reference**: Historical performance data, team strategies, and player statistics.
- **NFLverse Package**: For detailed NFL player and team data.
- **News and Social Media Data (Twitter, Reddit)**: Real-time team and player updates.
- **Video Data (YouTube, Game Footage)**: For player detection using YOLO and enhancing tracking analysis.
- **Player Mobility Data**: Using Next Gen Stats and historical tracking data to train models for player trajectory prediction.

Potential Approach & Technologies

- **YOLOv8** for real-time video analysis and player detection.
- **DeepSeek-7B** (fine-tuned for NFL-specific tasks) for interpreting NFL playbooks and team strategies.
- **GNN** for analysis of player spatial relationships.
- **LSTM/RNN** for predicting player mobility and trajectory.
- **Generative Transformers** (GPT, T5) for creating dynamic, real-time reports.
- **AWS Cloud Services** for seamless data integration, storage, and computational power (including AWS Neptune for GNN-based models).
- **Python** (TensorFlow, PyTorch) for developing and training AI and machine learning models.
- **RAG Framework** for dynamically integrating new information into AI predictions to ensure up-to-date insights.

Expected Deliverables

- **AI-Generated Scout Reports**
    - Pre-Game: Opponent strength/weakness assessment.
    - In-Game: Live adjustments based on real-time player tracking & sentiment shifts.
    - Post-Game: Evaluation and lessons for future matchups.
- **Interactive Dashboard**
    - Data visualizations of player movement, opponent strategies, and game conditions.
    - Real-time strategy recommendations.
- **Open-Source Code Repository**
    - Scripts for YOLO detection, pose analysis, sentiment extraction, and RAG-based insights.

Timeline & Milestones

Week 1-2: Data Collection & Preprocessing

- Gather data from nflverse, Pro Football Reference, Next Gen Stats, and social media.
- Preprocess video data for pose estimation.

Week 3-4: Model Development

- Train YOLO model for player tracking.
- Fine-tune DeepSeek-7B for NFL playbook insights.

Week 5-6: RAG & Multi-Modal Fusion

- Implement vector-based retrieval for real-time news & social media.
- Develop Graph Neural Networks (GNN) for player interaction modeling.

Week 7-8: Dashboard & Testing

- Deploy interactive dashboard with real-time strategy updates.
- Conduct pilot testing with historical game footage.