Formula 1 Race Penalty Analysis and Explanation

Lorraine Bichara Beatrice Chen Bhargav Kosaraju

Background and Motivation

Formula 1 is governed by the FIA (International Automobile Federation) that provides penalties to keep racing fair and safe for drivers. For an average fan, penalties can be confusing as they vary widely in type and impact, and are determined by the FIA Sporting and Technical Regulations which is a highly technical document. Because it's not clear why one penalty may receive a harsher sanction than another, it makes the system feel inconsistent and unpredictable.

Problem Statement

Develop an application that, given a race penalty, can provide a more user-friendly explanation of the infringement referencing the FIA sporting regulations, as well as an analysis of the fairness of the penalty received compared to past penalties.

Data Sources

- Source: Textual data, including official Sporting Regulations and steward decision documents in PDF format, will be sourced from the official FIA website [1]. Numerical data for historical race sessions will be retrieved using the Python package fastf1 [2].
- Description: Race control documents provide semi-structured text detailing penalties for each race incident with reference to the official sporting regulation clauses. Numerical data contains granular car performance data during each race session, including lap timing, car telemetry, session results, etc.
- Key Attributes: For textual data, key attributes include the race session name, identifier of the breached regulation clause, drivers involved, incident timestamps and the explanation text. For numerical data, the most relevant attributes are car telemetry (speed, throttle, brake, etc.), lap times, and positioning data.
- Relevance: The textual data provides the official legal basis and reasoning for penalties, which is essential for the accuracy of the explanation provided by the Large Language Model (LLM). The numerical telemetry data offers objective evidence of the on-track events, allowing the system to ground its explanations in verifiable facts.
- Data Quality Concerns: The fastf1 library relies on an unofficial API, which poses a risk of rate-limiting or future breaking changes.

Scope and Objectives

Race penalty explanation and fairness analysis via regulation-based reasoning + fan-facing educational support.

Objectives:

- Collect and preprocess historical F1 penalty data to build a structured dataset of penalties and race contexts

- Develop a natural language processing model to match penalties with relevant regulation references
- Implement a fairness analysis engine that compares current penalties against historical precedents by measuring consistency in sanctions for similar infractions. To mitigate LLM hallucinations, all explanations will be grounded in retrieved documents
- Design an intuitive and user-friendly interface for fans to interact by querying for penalties and displaying explanations.

Minimum Components for a Good Project

- Large Data: The project will process a large, multi-modal dataset, combining years of FIA regulatory PDFs with thousands of steward decision documents and high-frequency, structured numerical data from the fastf1 library.
- Scalability: The system is designed for scalability, using a containerized architecture that can be deployed on cloud infrastructure to handle simultaneous queries from many users.
- Complex Models: The project's core is a complex, agentic Retrieval-Augmented Generation (RAG) system. This system will use an LLM to decompose user queries, call corresponding functions to retrieve information from the backend databases, and use the returned information to form a detailed explanation for the race decisions. The base LLM (Gemini 2.5) will not be fine-tuned end-to-end. Instead, lightweight domain adaptation will be performed through prompt engineering using a curated dataset of FIA steward decision documents. The objective is to improve the model's ability to reference specific regulations accurately, explain penalties in accessible language, and evaluate the fairness based on historical precedents. This domain-specific augmentation will help with precision in the model's responses.
- Computationally Expensive Inference: Inference involves a multi-step process of data retrieval, re-ranking, and synthesis by the LLM. A key focus will be optimizing this pipeline to minimize latency, which is critical for real-time use cases like a live commentator assistant.

Learning Emphasis

The project will be implemented utilizing the MLOps content covered in the course. The application aims to leverage LLM with a RAG system for the core functionality. The deployment will use a containerized environment with CI/CD automation via GitHub Actions.

Application Mock Design

- UI Design Prototype (made using Figma): https://twine-claw-08738571.figma.site/
- Block Diagram: Mock Design Details

Research and Development

[1] "Decision documents," Federation Internationale de l'Automobile. Accessed: Sept. 25, 2025. [Online]. Available: https://www.fia.com/documents/championships/fia-formula-one-world-championship-14/season/season-2025-2071

[2] P. Schaefer, *theOehrly/Fast-F1*. (Sept. 25, 2025). Python. Accessed: Sept. 25, 2025. [Online]. Available: https://github.com/theOehrly/Fast-F1

Fun Factor

All three of us are fans of Formula One and enjoy following the sport as a hobby. Because of our technical background, we're also interested in how they're leveraging data to make decisions, so we're excited to explore this space.

Milestones

1. Data collection and preprocessing - 10/6
2. Environment set up - 10/6
3. RAG implementation - 10/14
4. *Deliverable: Milestone 2 - 10/16*
5. Mid-term presentation - 10/28
6. *Deliverable: Milestone 3 - 10/28*
7. Frontend development - 11/13
8. Continuous Integration & Testing - 11/20
9. Final testing and deployment - 11/24
10. *Deliverable: Milestone 4 - 11/25*
11. Video presentation - 12/8
12. *Deliverable: Milestone 5 - 12/11*