WHITEPAPER

Unlock Your Streaming's True Potential Now

Advancing Streaming Piracy Detection and Prevention

Our cutting-edge anti-piracy solution leverages AI, scalable data and cloud engineering, and user-centric design thinking to safeguard content, deliver real-time insights, and effectively enhance user experience in combating piracy.

fractal

1. Introduction

1.1 INDUSTRY LANDSCAPE

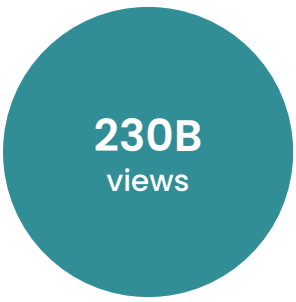
In the Media & Entertainment industry, combatting content piracy, copyright infringement, and intellectual property theft has proven to be an ongoing and evolving challenge. Despite substantial investments in legal and technological measures, effectively addressing these malicious practices remains daunting. While the advent of streaming platforms initially showed promise and resulted in a temporary decline, recent years have witnessed a troubling resurgence of digital piracy.

This resurgence can be attributed to various factors. The insatiable consumer demand for immediate access to exclusive content, coupled with economic and political instability, shifts in viewing habits during COVID-19, and the unregulated proliferation of generative AI technology, collectively contribute to piracy's prevailing and expanding presence. It is essential to recognize that piracy undermines the revenue streams of content producers and artists and exposes consumers to significant risks. These risks encompass malware distribution, identity theft, counterfeit product dissemination, and potential ties to terrorism.

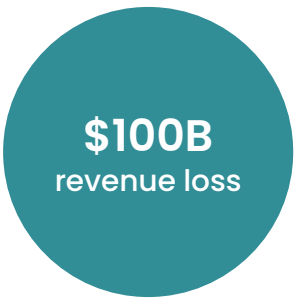
Addressing this issue requires a multifaceted approach. It involves leveraging legal frameworks, deploying advanced technological solutions, and promoting consumer education and awareness. Industry stakeholders must collaborate closely to develop and enforce robust anti-piracy measures. Additionally, exploring innovative business models, enhancing content accessibility, and ensuring a seamless user experience can help mitigate the appeal of piracy.

By tackling content piracy head-on, the Media & Entertainment industry can safeguard the interests of content creators, protect consumers from potential harm, and foster a thriving and sustainable digital ecosystem.

The industry's current shortcomings and failure are quite evident in the alarming statistics*:



Pirated streaming contents receive upwards of 230B views per year



Media industry loses close to $100 billion annually due to digital privacy

Around 500k jobs are lost in the entertainment industry every year due to piracy

**Global statistics based on secondary research (FY'22-23)*

1.2 PROBLEM FRAMING

In the face of mounting concerns surrounding their platform, a leading video-streaming service provider recognized the need for an intelligent solution to combat piracy proactively. With millions of daily active users spanning diverse geographical regions, and a vast library encompassing live sports, movies, TV series, and more, their dedicated anti-piracy team encountered numerous critical challenges:

- **Uncertainty regarding typical pirate behavior**
- **Inability to pinpoint users at elevated risk of piracy**
- **Dependence on manual and subjective decision-making procedures**
- **Limited visibility into piracy catalysts and indicators**
- **Inability to adapt to evolving piracy patterns**

To overcome these challenges, the streaming provider joined forces with Fractal to deconstruct the overarching problem into smaller objectives. We aimed to create a robust anti-piracy framework by leveraging our expertise in AI. The following requirements were identified for the framework:

AUTOMATED RISK SCORING	Automated risk scoring of accounts, based on streaming behaviour, IP address, device usage and other attributes.
BEHAVIOR IDENTIFICATION	Identification of other accounts with attributes and behaviours similar to identified pirates.
IP ADDRESS / DEVICE TRACKING	Near real-time monitoring of IP addresses and devices being used across multiple accounts and geographic regions.
DEVICE USAGE MONITORING	Investigating suspicious devices to identify how those are used across accounts, to terminate or suspend the account.
DEVICE USAGE MONITORING	Investigating suspicious devices to identify how those are used across accounts, to terminate or suspend the account.

2. Anti-piracy framework

2.1 KEY PRINCIPLES

Fractal has developed an approach rooted in three essential pillars of problem-solving to devise a cutting-edge anti-piracy framework that fulfills the identified requirements.

DATA & TECH

Capturing underlying dynamic user behaviour at each point in the event journeys through real-time processing of millions of data-points

AI & MACHINE LEARNING

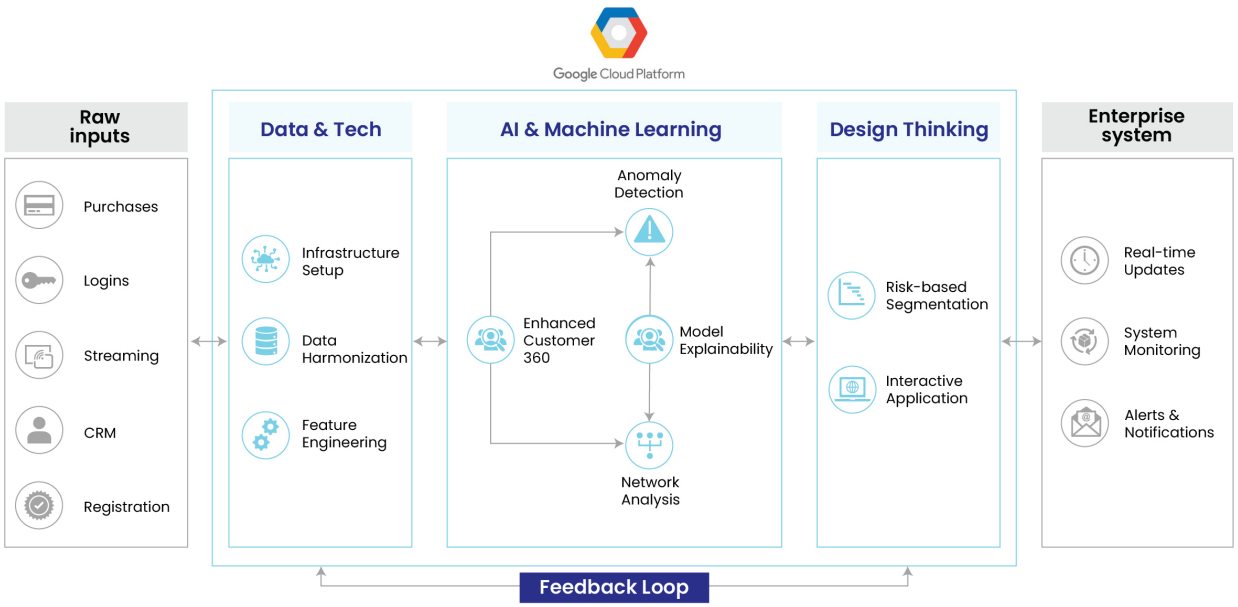
Building a robust and transparent algorithmic solution which can learn in real-time, identify potential pirates and drive actionable insights

DESIGN THINKING

Designing an easy-to-use interactive Web UI to support real-time insight generation, self-service and manual over-rides, if required

2.2. FRAMEWORK SUMMARY

The foundation of the framework was established on the robust infrastructure of Google Cloud Platform (GCP), but the principles and open-source codebases can be readily expanded to other public or private cloud environments (such as AWS, Azure) or on-premises setups, with minimal adjustments. Now, let's explore each key component to gain a comprehensive understanding.



Engineering:

Cloud and data engineering forms the foundation of the architectural design, enabling enhanced efficiency, scalability, and insights, and comprises the following components:

- **Infrastructure setup**

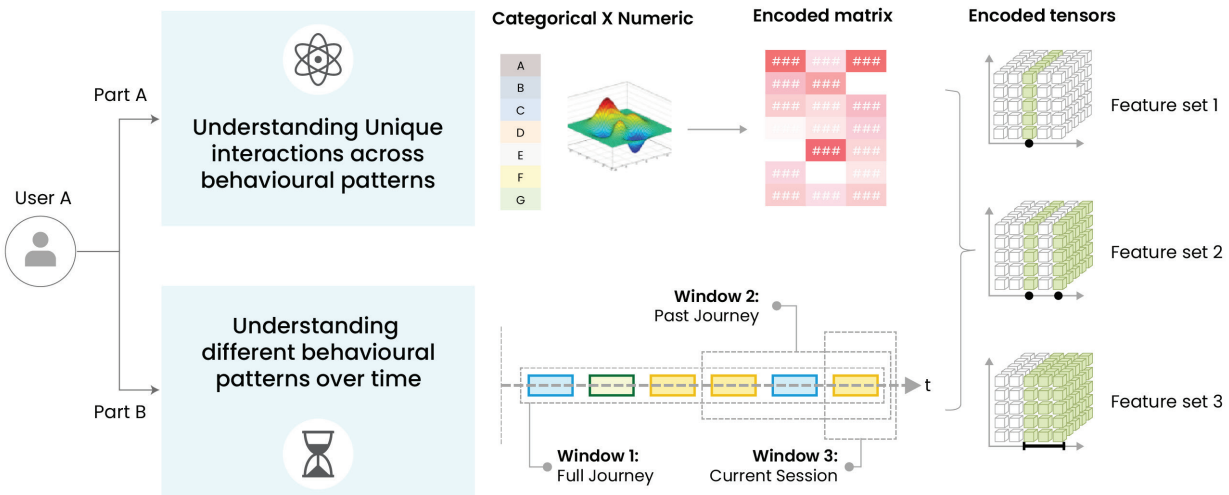
To establish a scalable computational ecosystem, we leveraged cloud-native services like BigQuery, Compute Engine, Cloud Storage, GCP Marketplace (Neo4J), and Google AppSheet. Equivalent services are available on other popular cloud service providers such as AWS and Azure. For data processing, we utilized Python libraries like Pandas and NumPy, enabling both standard and customized data processing capabilities.

- **Data harmonization**

We consolidated more than **10 disparate data sources,** encompassing past purchases, logins, streaming sessions, demographics, and registrations. Through this process, we established a single source of truth known as **Customer 360**. This unified data repository enables us to generate a comprehensive user context, empowering us with deep insights into user behavior, preferences, and engagement patterns.

- **Feature engineering**

We employed advanced data transformation and feature creation techniques to curate an **enhanced Customer 360,** comprising over **500 features**. These features are the foundation for quantifying customer behavior across multiple dimensions and capturing diverse latent interactions.



Artificial Intelligence:

The intelligent algorithm layer at the core of the anti-piracy framework comprises three key components:

- **Anomaly detection**

Building upon the Enhanced Customer 360 data set, we leveraged unsupervised auto-tuning anomaly detection algorithms to evaluate user behavior at the individual session/stream level. This approach enables continuous self-learning and adaptation, ensuring our system remains vigilant against evolving piracy trends without the need for manual intervention. Mathematically, the model can be summarized as follows:

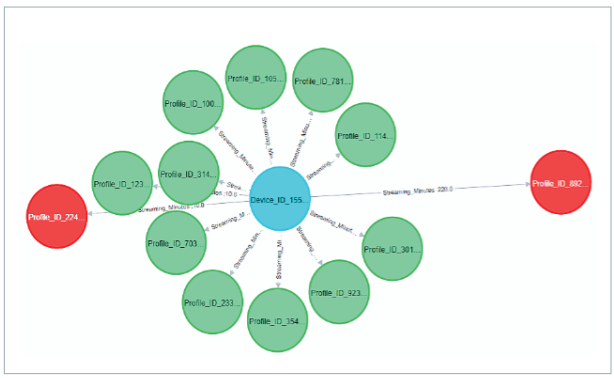
$$r\ (u, s) = f\ (\text{Feature set 1, Feature set 2, Feature set 3}) \longrightarrow \boldsymbol{R\ [0,100]}$$

where r (u, s) = risk score for user "u" during session "s"; f = anomaly detection model; Feature set 1–3 = features from Enhanced Customer 360 data set; R [0,100] = any real number between 0-100.

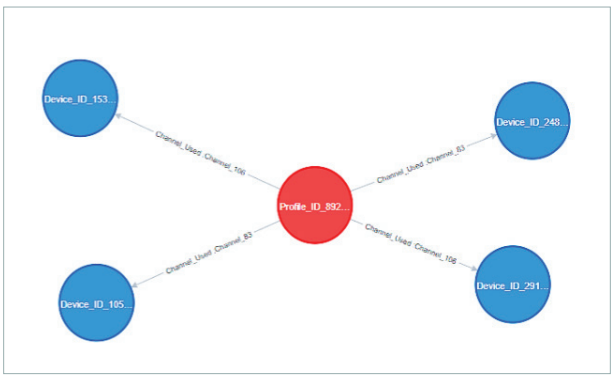
While standalone anomaly detection plays a vital role, it is crucial to recognize that its effectiveness can be further enhanced through a holistic approach. Combining network analysis techniques and model explainability methods can significantly improve piracy identification's precision while minimizing false positives.

- **Network analysis**

We constructed dynamic graphical networks that trace extensive syndicates of users, devices, and IP addresses linked to high-risk individuals identified by the anomaly detection framework. To illustrate the impact and application of this approach, consider these examples that offer a clear perspective on its effectiveness.



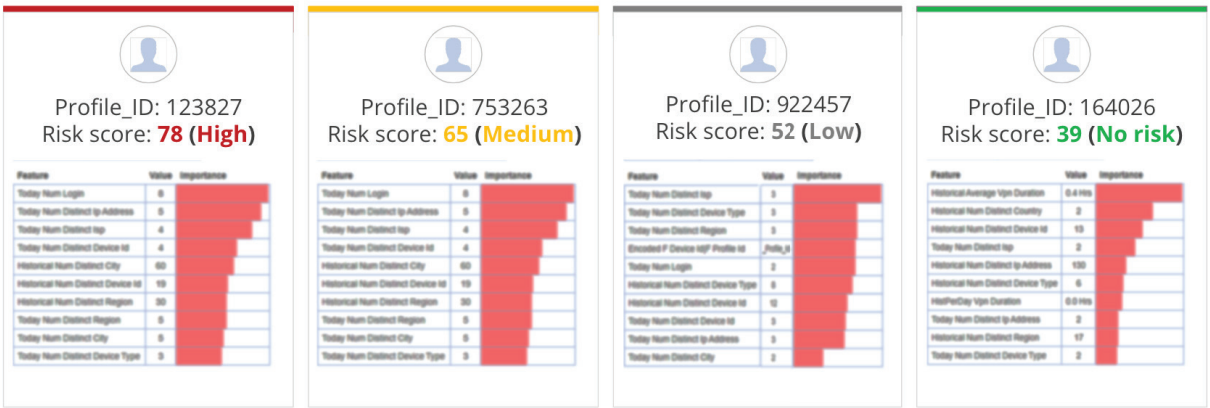
Example 1: Specific **device** mapped to identified **pirates** and at-risk **non-pirates.**



Example 2: Particular **pirate** mapped to multiple **devices** concurrently.

- **Model explainability**

We prioritized transparency by offering comprehensive visibility into the inner workings of our ML models. By providing in-depth insights into crucial indicators and drivers, the system allows users to validate, investigate, and take specific actions based on observed patterns.
This functionality goes beyond mere risk scores, enabling users to understand the underlying factors contributing to piracy.

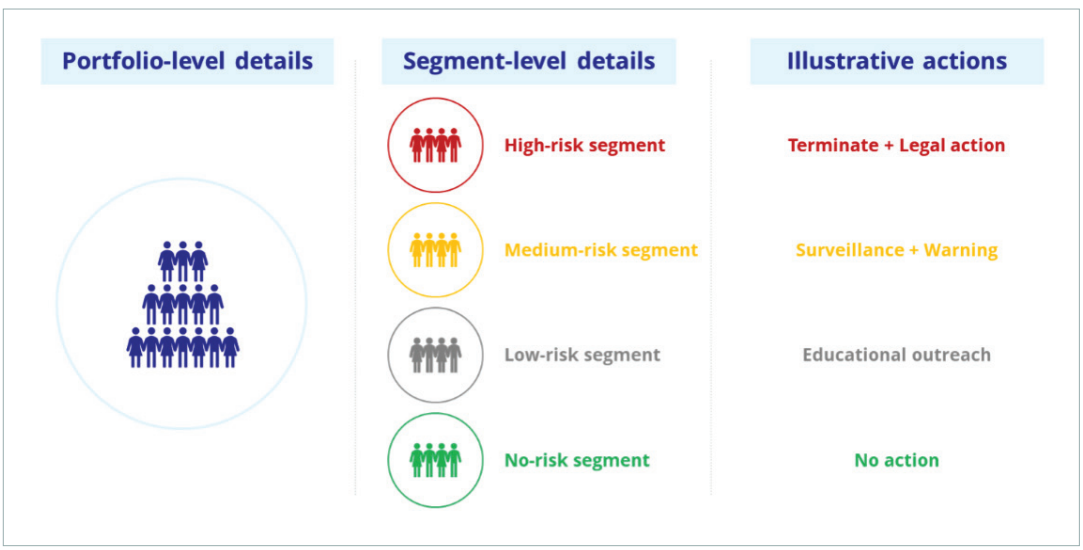


Design thinking:

Building upon the principles of transparency, this module places a strong emphasis on engaging human users in decision-making processes and tailoring interventions. It encompasses two critical components that facilitate this approach:

- **Risk-based segmentation**

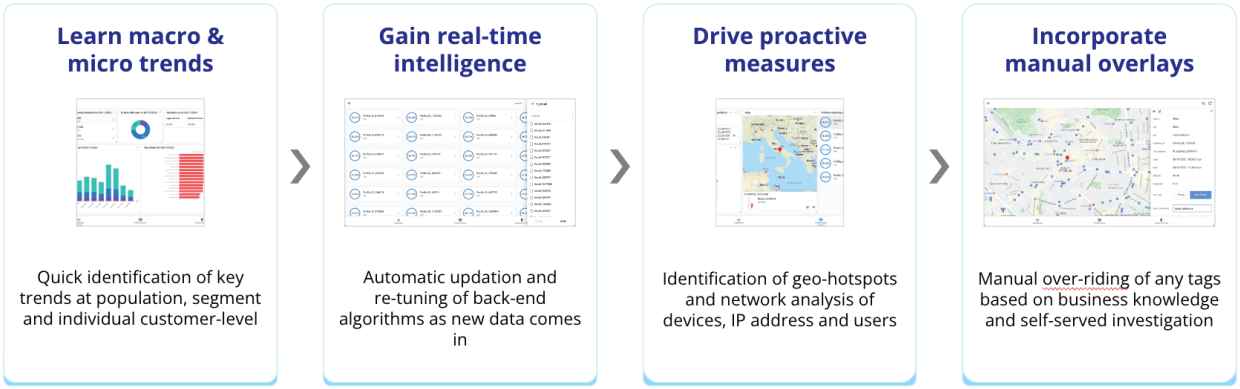
By segmenting users based on their risk levels, we tailored the actions and interventions for each segment. This targeted approach addressed the challenge of personalizing interventions at the user level, while also enabling broader cohort-level interventions, which are highly effective.



- **Interactive Application**

To facilitate thorough investigations and incorporate human judgment in sensitive scenarios, we provided piracy analysts with a self-serve web interface powered by Google AppSheet. This "human-in-the-loop" design allows analysts to delve deeper into cases and override algorithmic decisions when necessary. The interface offers key functionalities essential to their work, including:

- Comprehensive summary reports on piracy events
- Identification of major geographical hotspots for piracy activities
- Network analysis revealing connections between users, IP addresses, and devices involved in piracy
- In-depth driver analysis of piracy events, unraveling the underlying factors
- Trend analysis to uncover emerging piracy behaviors and patterns.



3. The impact

Quantifying the monetary value of piracy reduction resulting from our framework requires a comprehensive analysis involving counterfactual evaluation, quasi-experimentation, and market research. While this topic deserves a separate discussion beyond the scope of this article, we can still provide a high-level understanding of the business value through the following:

- **Reduction in revenue loss**

The framework aims to minimize revenue loss resulting from piracy by actively combating illicit activities and protecting content creators' revenue streams.

- **Safeguarding brand reputation**

By mitigating piracy, the framework helps maintain the organizations' brand reputation, fostering trust among consumers and content creators.

- **Understanding pirates**

With the capability to process over 1 million streaming sessions daily, our framework identified approximately 0.5–1% of sessions associated with potential pirates. Analyzing key indicators allows us to create a data-driven profile of a typical pirate across more than 100 dimensions. To illustrate this, consider the image below, which highlights three such dimensions: VPN usage, streaming duration, and device usage:

Pirates accessed VPN more often than non-pirate users

High-risk users streamed continuously for longer durations

Piracy users accessed more devices than non-pirates

- **Accelerating data-to-decision time**

Our real-time intelligence framework significantly reduces the time required for analysts to identify, track, and investigate suspicious accounts. With this advanced system in place, the data-to-decision time is streamlined by an impressive 60–80%. Previously, these tasks would typically take weeks or even months to complete, but with our framework, they can now be accomplished within a matter of days.

- **Minimizing errors**

Manual tasks and laborious processes are susceptible to errors, which can lead to substantial business and legal implications. Our framework addresses this challenge by implementing end-to-end automation, significantly reducing the potential for human error. While algorithms handle most cases, a few critical situations are escalated to end users for their expertise and input.

4. Authors

Daryna Mozolova
Associate

Oleksii Galganov
Associate

Oleksii Panichek
Associate

Yuvaneet Kumar Bhaker
Principal Data Scientist

Chandramauli Chaudhuri
Principal Data Scientist

Sumit Tayal
Client Partner

5. References

- Business Wire
- Forbes
- Statista
- The Global Innovation Policy Center
- NERA Economic Consulting
- Motion Picture Association

About Fractal

Fractal is one of the most prominent providers of Artificial Intelligence to Fortune 500® companies. Fractal's vision is to power every human decision in the enterprise, and bring AI, engineering, and design to help the world's most admired companies.

Fractal's businesses include Crux Intelligence (AI driven business intelligence), Eugenie.ai (AI for sustainability), Asper.ai (AI for revenue growth management) and Senseforth.ai (conversational AI for sales and customer service). Fractal incubated Qure.ai, a leading player in healthcare AI for detecting Tuberculosis and Lung cancer.

Fractal currently has 4000+ employees across 16 global locations, including the United States, UK, Ukraine, India, Singapore, and Australia. Fractal has been recognized as 'Great Workplace' and 'India's Best Workplaces for Women' in the top 100 (large) category by The Great Place to Work® Institute; featured as a leader in Customer Analytics Service Providers Wave™ 2021, Computer Vision Consultancies Wave™ 2020 & Specialized Insights Service Providers Wave™ 2020 by Forrester Research Inc., a leader in Analytics & AI Services Specialists Peak Matrix 2022 by Everest Group and recognized as an 'Honorable Vendor' in 2022 Magic Quadrant™ for data & analytics by Gartner Inc.

For more information, visit fractal.ai

Corporate Headquarters
Suite 76J,
One World Trade Center, New York,
NY 10007

Get in touch