Data Visualization Project

Bot Detection Dashboard for Social Media Accounts

Fahim Shahriar
University of Southern Denmark
Odense, Denmark
fasha23@student.sdu.dk

Adib Abzaal
University of Southern Denmark
Odense, Denmark
adabz23@student.sdu.dk

December 26, 2025

Percent of Contribution (PoC)

Percent of Contribution for the Group Project

Student	SDU ID / E-mail	Percent of Contribution (%)
Fahim Shahriar	fasha23@student.sdu.dk	100
Adib Abzaal	adabz23@student.sdu.dk	100

Abstract

This project presents an interactive data visualization dashboard for analysing behavioural differences between bot and human accounts in a social media dataset. The dashboard combines static, animated, and interactive visualizations to explore profile characteristics, activity levels, engagement behaviour, and content features. The results reveal clear distinctions between bots and humans, including more extreme follower–following behaviour, higher posting activity, and lower engagement for bot accounts. The project demonstrates how visual analytics, supported by interactive storytelling and AI-assisted summaries, can make complex bot detection patterns more transparent and accessible to non-expert users.

1 Background and Motivation

Now a days social media is a common platform for communication, and people use it for everything from daily updates to public discussions. However, Automated accounts, usually called *bots*, are common as well. Some bots are harmless (posting routine updates), but others are built to manipulate activity—by boosting certain posts, spamming URLs, or making engagement look higher than it is. In this project, we therefore focus on the practical question: can we see consistent behavioural differences between bots and humans in the data, using observable features rather than guessing based on a single metric?

A common approach to bot detection is to rely on machine learning models that classify accounts based on behavioural features. While these models can achieve strong predictive performance, they often behave like black boxes: they provide a final label (bot or human) but do not clearly explain *why* an account was flagged. This lack of transparency can reduce trust in the results, especially for non-expert users such as moderators, journalists, researchers from other fields, or everyday platform users. As a result, there is a practical need for methods that complement automated classification with interpretable evidence and exploratory insight.

Data visualization is well suited to address this gap. By converting behavioural features into visual patterns, users can compare bot and human behaviour directly and build intuition about what distinguishes the two groups. For example, differences in follower–following ratios, posting frequency, and engagement levels become easier to interpret when displayed through boxplots, histograms, and comparative bar charts. Content-related features such as URL usage, text variability, and emotion cues can be explored through scatter plots and feature comparisons. Temporal features can also reveal whether activity follows a regular automated pattern or a more irregular human routine. In short, visualization helps users move from abstract numbers to concrete evidence.

The motivation for this project is therefore twofold. First, we aim to investigate behavioural differences between bot and human accounts across multiple dimensions, including profile and popularity signals, activity and engagement patterns, content style, and temporal behaviour. Second, we aim to demonstrate how an interactive dashboard can support deeper understanding than static plots alone. By allowing users to explore the data through multiple coordinated views—combined with animation and AI-assisted summary graphics—the dashboard supports both overview-level understanding and detailed exploration. Ultimately, the project contributes a human-centred visual analytics approach that makes bot detection patterns more transparent, interpretable, and accessible.

2 Project Objectives

List the research questions: The main objective of this project is to explore and communicate behavioural differences between bot and human social media accounts through effective data visualization and visual analytics techniques. Rather than focusing solely on automated classification, the project aims to provide interpretable, interactive, and visually driven insights that help users understand how and why bots differ from genuine human users across multiple dimensions.

To achieve this goal the project is structured around a set of research questions that guide the design of the dashboard and the selection of the visualizations. These objectives focus on analysing account popularity, activity patterns, engagement behaviour, content characteristics and temporal dynamics using both static and animated visual representations.

Specifically the project seeks to achieve the following objectives:

- **To analyse the profile-level and popularity related differences between bot and human accounts** with particular focus on follower-following behaviour, profile completeness indicators and user rant distributions (**RQ**1). This objective aims to determine whether popularity-based features provide meaningful signals for distinguishing automated accounts from human users.

- **To examine differences in activity and engagement patterns between bots and humans**, including posting frequency and received engagement in terms of likes, comments, and reposts (**RQ**2). The goal is to understand whether higher activity levels translate into higher engagement or whether bots exhibit fundamentally different interaction dynamics.
- **To investigate variations in content style between bot and human accounts**, focusing on URL usage, the text length variability, punctuation, and emotion-related features (**RQ**3). This objective explores whether linguistic and content-based characteristics reveal systematic differences in how automated and human accounts generate posts.
- **To explore the temporal posting behaviour of bots and human accounts**. including average posting times, variability and time-related parameters (**RQ**4). The aim is to identify whether bots exhibit more regular and predictable temporal patterns compared to human users.

The list of research questions is given below:

1. **RQ1:** How do bots differ from humans in follower–following behaviour, profile completeness, and user rank?
2. **RQ2:** Do bots and humans differ in posting frequency and the engagement they receive (comments, reposts, likes)?
3. **RQ3:** Are there differences in URL usage, text length, punctuation, and emotion usage between bots and humans?
4. **RQ4:** Are there differences in posting time behaviour (average posting time, variability, and time-related parameters) between bots and humans?

In addition to addressing these analytical objectives, the project also aims to demonstrate the effectiveness of interactive and animated visualisations in supporting exploratory data analysis. By integrating multiple visualisation types, animation, and AI-assisted visual analytics into a single dashboard, the project seeks to provide a comprehensive and user-friendly tool that enables both researchers and non-technical users to explore and interpret complex social media behaviour.

3 Data

3.1 Data Source

The dataset used in this project is the SocialBot dataset, provided as part of the course material of another course which we took in the last semester. This dataset is also approved by the professor of this course. The data consists of account-level information collected from a social media platform and is designed to support the analysis of behavioural differences between bot and human accounts. The dataset is distributed is spreadsheet format(socialBot.Xlsx) and contains both numerical and categorical variables describing user profiles, activity patterns, engagement metrics and content-related features.

The dataset includes a binary label indicating whether an account is classified as a bot or a human. This label serves as the ground truth for comparative analysis throughout the project. Since the dataset is pre-collected and anonymised, no personal or sensitive user information is exposed, making it suitable for academic analysis and visualisation purposes.
The SocialBot dataset was selected because it provides a rich and diverse set of features that allow the exploration of multiple behavioural dimensions, including popularity, engagement, content style, and temporal activity. This makes it particularly well suited for a visualization-driven investigation of bot detection patterns.

3.2 Data Description

The socialBot dataset contains records representing individual social media accounts, with each record describing aggregated behavioural characteristics of an account. The dataset includes a mixture of numerical and categorical variables capturing different aspects of account behaviour. Key variables used in this project include:

- Account type(is_bot): A binary indicator identifying whether an account is a bot or a human. For visualisation clarity, this variable is treated as a categorical factor with labels Bot and Human.
- Follower-following rate(follower_following_rate): A ratio-based feature representing the relationship between the number of followers and the number of accounts followed. This variable is used as an indicator of popularity and growth behaviour.
- User rank (urank): A numerical measure representing the relative ranking or popularity of an account within the platform.
- Activity-related features: Variables describing posting behaviour such as posting frequency and average activity levels.
- Engagement metrics: Aggregated measures of user interaction, including average likes, comments, and reposts received per post.
- Content-related features: Variables capturing content style and structure, such as URL usage, text length variability, punctuation patterns, and emotion-related token usage.
- Temporal features: Time-based variables describing posting behaviour, including average posting time and variability over time.

The dataset contains a sufficient number of observations to support meaningful visual comparisons between bot and human accounts. variables vary in scale and distribution, ranging from ratio-based metrics to count-based and normalised features. This diversity makes the dataset well-suited for multi-dimensional visualisation techniques such as boxplots, scatter plots, histograms, correlation heatmaps and principal component analysis.
Overall, the dataset provides a comprehensive foundation for exploring behavioural differences between automated and human-operated social media accounts through visual analytics.

3.3 Data Processing

Before creating the visualizations, several data processing steps were applied to ensure that the dataset was suitable for visual analysis and comparison between bot and human accounts. The original dataset (`SocialBot.xlsx`) contains account-level features related to profile characteristics, activity patterns, engagement metrics, and content properties.

First, categorical variables were cleaned and standardized. The variable `is_bot`, originally encoded as binary values, was converted into a categorical factor with meaningful labels (*Human* and *Bot*) to improve readability in the visualizations. In addition, user rank values (`urank`) were discretized into three rank groups (*Low*, *Medium*, and *High*) to support grouped comparisons and animated storytelling.

Missing values were handled conservatively to preserve as much data as possible. For aggregated statistics such as mean engagement values (likes, reposts, and comments), missing observations were ignored using pairwise deletion. For multivariate analyses such as correlation matrices and principal component analysis (PCA), only complete numeric records were used to ensure numerical stability and avoid biased results.

Several derived features were created to better capture behavioural differences between bots and humans. These include follower–following ratios, posting rate summaries, engagement averages, URL usage statistics, text length variability, and emotion-related content features. For visualization purposes, selected variables were reshaped into long format, enabling the use of faceted plots and grouped bar charts.

Lastly, to make sure that variables with different scales contributed equally to the results, numerical features were normalized where necessary, especially for PCA and clustering analyses. Maintaining transparency, reducing needless data loss, and preparing the dataset to enable understandable visual comparisons between bot and human accounts were the main goals of the data processing.

4 Visualization and Dashboard

4.1 Design Choices

The dashboard was designed with the goal of supporting visual analysis while remaining user friendly design. To achieve this, the interface follows a clear and structured layout where different analytical perspectives are separated into tabs with icons. Each tab corresponds to a specific research question or analysis goal, allowing users to gradually move from an overview of the dataset to more detailed and interactive explorations.

A sidebar-based navigation layout was chosen to provide consistent and intuitive access to all dashboard components. The main tabs include *Overview*, *Profile & Popularity*, *Activity & Engagement*, *Content & Timing*, *Principal Component Analysis (PCA)*, *Interactive Story*, and *AI-generated Graph*. This organisation helps users maintain context and reduces cognitive load by grouping related visualizations together.

Visual encoding choices were made to ensure clarity and consistency across the dashboard. Account type is encoded using colour, with human accounts shown in green or blue and bot accounts shown in red, making comparisons immediately visible across all plots. Quantitative variables such as follower–following rate, posting rate, engagement metrics, and content features are mapped to position and scale, which are known to be among the most accurate visual encodings for numerical data. Where appropriate, faceting is used to display multiple related measures within a single figure without overcrowding the visual space.

Interactivity plays a central role in the dashboard design. Users can explore patterns through hover tooltips, dynamic filters, and animated transitions. In particular, the interactive story tab allows users to examine how engagement behaviour changes under different follower–following rate scenarios, while the decision boundary and clustering tools support deeper analytical exploration. Animation is used selectively to highlight behavioural changes over stages rather than as a decorative element.

4.2 Features

The dashboard was designed and implemented to fully satisfy all mandatory requirements specified in the course project guidelines. Each must-have feature was integrated in a way that supports both analytical exploration and clear communication of results.

At first, the dashboard includes a wide range of visualization types in order to capture different aspects of the data. These include bar charts, boxplots, scatter plots, histograms, heatmaps, and parallel coordinate plots. For instance, boxplots are used to compare follower–following ratios and user rank between human and bot accounts (see figs. 1 and 2), while histograms provide insight into the distribution of follower–following rates across account types (see fig. 3). Scatter plots and PCA-based visualizations further support multivariate analysis (see figs. 6 and 8).

Second, at least one animated visualization is included to illustrate how behavioural patterns change under different scenarios. The animated story focuses on engagement behaviour across varying follower–following rate levels, allowing users to observe how average likes shift dynamically rather than relying on static comparisons. A representative frame of this animation is included in the appendix (see fig. 14).

Third, an AI-generated visualization is incorporated to provide a high-level summary of relationships between numeric features. Specifically, an AI-assisted correlation heatmap is used to highlight global associations among behavioural, engagement, and content-related variables. This visualization supports exploratory reasoning by revealing patterns that may not be immediately visible in individual plots (see fig. 9).

So, the dashboard contains well over the required nine visualizations, distributed across thematic tabs such as Overview, Profile & Popularity, Activity & Engagement, Content & Timing, PCA, and the Interactive Story. Several interactive analytical tools are also provided, including a clustering explorer, a decision boundary visualisation, and a bot detection sandbox, which allow users to experiment with feature values and observe model-driven responses (see figs. 11 to 13). Additionally, the dashboard has the ability to export documentation and results.

4.3 Special Features

The Interactive Story, an important optional feature of the dashboard, allows actively investigate behavioral patterns and model behavior, going beyond conventional static visualizations. This section uses model-based visual analytics, animation, and interaction to facilitate exploratory learning and interpretation. The tab contains following story tabs:

- The *Animated Story* tab presents an animated scatter plot that illustrates how engagement changes under different follower–following rate scenarios. By stepping through predefined stages, observe how average likes vary across account types and rank groups, making behavioural differences easier to interpret over changing conditions (see fig. 14).

- The *Decision Boundary* tab visualises the output of a logistic regression model by displaying predicted bot probabilities as a background surface, with actual accounts overlaid as points. This helps understand how the model separates bots from humans across different feature combinations and supports transparency in model-driven analysis (see fig. 13).
- The *Clustering Explorer* tab enables unsupervised exploration using k-means clustering. By adjusting the number of clusters and selected features to observe how accounts group together and how bot and human accounts are distributed across clusters (see fig. 12).
- The *Detection Sandbox* provides a parallel coordinates view in which define a hypothetical account by adjusting feature values. The resulting scenario is visualised alongside real accounts, allows intuitively compare multivariate behaviour and examine how different feature settings relate to bot detection (see fig. 11).
- The *Feature Importance* tab displays the relative importance of features derived from a logistic regression model. This view supports interpretability by showing which variables contribute most strongly to distinguishing bot accounts from human accounts (see fig. 10).

4.4 Dashboard Link and Report Download

The interactive dashboard developed for this project is publicly accessible online and allows users to explore all visualizations, animations, and interactive analysis components in real time. A permanent link to the deployed dashboard is provided below.

In addition to the live dashboard, a report download feature is integrated directly into the interface to support documentation and offline review.

The project contains two download functionality:

1. *Analysis report* Automatic generation of a PDF report that summarises the dataset, visualizations, and key analytical insights.
2. *Project report* manually download the project report from the dashboard where project visualization analysis are briefly described.

5 Story and Results

This section presents the main findings of the project by answering the research questions introduced earlier (see **RQ1**, **RQ2**, **RQ3**, **RQ4**). The results are derived from the visual patterns observed across the dashboard and are supported by both static and interactive visualizations. Where appropriate, additional figures included in the appendix are referenced to provide further evidence.

5.1 RQ1 – Profile and Popularity

Research Question 1 examines how bot accounts differ from human accounts in terms of profile-related characteristics and popularity indicators, particularly follower–following behaviour and user rank.

Dashboard URL:
https://019b1f12-b9c4-4abe-b67a-8df144c084fd.share.connect.posit.cloud/

The follower–following ratio comparison is shown using a boxplot, where the x-axis represents the account type (Human or Bot) and the y-axis represents the follower–following rate (see fig. 1). The boxplot illustrates the distribution, median, and variability of this ratio for each group. Bot accounts exhibit a wider spread and more extreme values compared to human accounts, indicating less balanced follower–following behaviour. Human accounts, in contrast, tend to cluster around moderate ratios, suggesting more organic social connections.

User popularity is further analysed using a boxplot of user rank (`urank`), where the x-axis again denotes account type and the y-axis represents rank values (see fig. 2). The visualization shows that bot accounts are more concentrated at lower rank values, while human accounts generally occupy higher and more stable rank ranges. This pattern suggests that bots are less successful in achieving sustained popularity within the platform.

These findings are supported by the PCA scatter plot, which projects multiple profile-related features into a lower-dimensional space. In the PCA visualization, the x-axis (PC1) and y-axis (PC2) represent the two principal components that capture the largest variance in the data (see fig. 8). Bot and human accounts form partially separable clusters, indicating that profile and popularity features contribute meaningfully to distinguishing between the two groups.

5.2 RQ2 – Activity and Engagement

Research Question 2 focuses on differences in posting behaviour and engagement outcomes between bot and human accounts.

Posting activity is visualised using a violin plot, where the x-axis represents account type and the y-axis represents posting rate (see the Activity & Engagement tab in the dashboard). The violin plot shows the full distribution of posting frequency, revealing that bot accounts tend to post more frequently and with greater variability than human accounts. Human accounts generally display lower and more consistent posting rates, which aligns with typical user behaviour.

Engagement metrics are compared using grouped bar charts, where the x-axis represents different engagement measures (average likes, comments, and reposts) and the y-axis shows their mean values for each account type. The results indicate that human accounts consistently receive higher engagement across all metrics, while bot accounts receive lower engagement despite higher posting activity. This suggests that increased activity does not necessarily lead to meaningful interaction for bots.

The histogram of follower–following rates (see fig. 3) further contextualises these findings by showing how extreme follower–following ratios are more common among bot accounts. This reinforces the observation that bots often rely on aggressive activity strategies that do not translate into audience engagement.

As with RQ1, PCA provides additional insight into activity and engagement patterns. In the PCA scatter plot, accounts with similar activity and engagement characteristics tend to cluster together, with bot and human accounts showing partial separation. This supports the conclusion that activity and engagement features jointly contribute to distinguishing behavioural differences between bots and humans.

5.3 RQ3 – Content Style

Research question 3 examines whether bot and human accounts differ in terms of content style and linguistic characteristics. This analysis focuses on features related to URL usage, text length variablity, punctuation patterns, and emotion-related tokens, as these attributes provide insight into how content is generated and structures by automated versus human-operated accounts.
The visualizations reveal clear and consistent differences in URL-related behaviour. As hsowin in Figure 4 bot accounts exhibit higher average URL usage and greater URL variability compared to human accounts. This pattern suggests that bots are more likely to include links in their posts, which is consistent with automated strategies aimed at content promotion, redirection , or information dissemination.Human accounts, in contrast tend to nclude URLs less frequently and with lower variablity, reflecting more organic posting behaviour.
Differences are also evident in text length and linguistic variability. Bot accounts display more uniform text length patterns, whereas human accounts show greater variability in the length of their posts. This observation indicates that human-generated content is more diverse and context-dependent while bot-generated content often follows repetitive or templated structures. Such regularity in text length is a common characteristic of automated content generation.
Further insight is provided by the relationship between average words per post and emotion-related tokens, illustrated in Figure 5. Human accounts generally demonstrate higher levels of emotional expression relative to their text length, while bot accounts tend to use fewer emotion-related tokens. This suggests that human users are more likely to convey sentiment, opinions or affective cues in their posts whereas bots typically produce more neutral or informational content.
Overall, the content style analysis shows that bots and human differ not only in how frequently they post, but also in how they coomunicate. URL-heavy, linguistically uniform, and emotionally sparse content is more strongly associated with bot accounts, while human accounts tend to produce more varied, expressive, and context-rich posts. These findings indicate that content-based features provide valuable complementary signals to activity and popularity metrics when distinguishing between automated and human social media behaviour.

5.4 RQ4 – Temporal Patterns

Research Question 4 focuses on whether bot and human accounts differ in their posting-time behaviour, including average posting time and how variable (or regular) the posting schedule appears. Our initial expectation was that bots would show more regular and predictable timing patterns, while humans would appear more irregular because their activity is shaped by daily routines, context, and spontaneity.

In our dashboard, temporal variables are part of the dataset's numeric feature set and are therefore reflected most clearly in the multivariate views rather than in a single standalone "time-of-day" chart. The first supporting view is the AI-generated correlation heatmap (Figure 9). Technically, this plot places numeric features on both axes and encodes their pairwise correlation using a diverging colour scale (stronger colour intensity indicates stronger correlation, with the midpoint around zero). By inspecting where time-related features fall in this matrix, we can see whether timing metrics behave like independent signals or whether they move together with other behavioural dimensions such as activity (posting rate) and engagement. Overall, the heatmap provides evidence that temporal behaviour should be interpreted as part of a broader behavioural signature, not as an isolated feature.

A second piece of evidence comes from the PCA scatter plot (Figure 8). In this visualization, each point is an account projected onto the first two principal components (PC1 and PC2), which summarize variance across many numeric variables at once. If temporal patterns were extremely distinctive on their own, we would expect a near-perfect separation between bots and humans driven primarily by time-related variance. Instead, the plot shows only partial separation, which suggests that temporal features contribute to distinguishing bots from humans, but mainly in combination with profile, activity, engagement, and content features rather than acting as a single dominant separator.

Finally, although it is not a direct "posting timestamp" plot, the animated story (Figure 14) supports the interpretation that behavioural differences become easier to see when we view changes across stages rather than relying on a single static snapshot. This storytelling approach reinforces the idea behind RQ4: regularity and predictability are best understood when behaviour is viewed dynamically and comparatively across conditions.

The temporal analysis indicates that bot and human accounts differ in posting-time behaviour, but the evidence from the heatmap and PCA suggests that timing signals are most informative when interpreted alongside other behavioural cues. A limitation is that our report primarily relies on account-level summary time features; future improvements could strengthen RQ4 by adding a dedicated temporal visualization (e.g., hour-of-day distributions, weekday vs weekend activity, or circular time plots) to show timing regularity more directly.

5.5 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is used to provide a compact, multivariate view of the dataset by reducing many correlated numerical features into a small number of uncorrelated components. This allows overall behavioural patterns of bot and human accounts to be examined beyond individual variables.

In the PCA scatter plot, the x-axis (PC1) and y-axis (PC2) represent the first two principal components, which capture the largest proportion of variance in the data (see fig. 8). Each point corresponds to an account, and colour indicates whether the account is labelled as Human or Bot. The visualization reveals partial separation between bot and human accounts, suggesting that a combination of profile, activity, engagement, and content-related features contributes to distinguishing the two groups.

Pairwise PCA-related scatter plots further support this observation. For example, the relationship between follower–following rate and URL variability highlights behavioural differences that align with findings from RQ1 and RQ3 (see fig. 6), while the comparison between follower–following rate and user rank reinforces popularity-related patterns discussed in RQ1 (see fig. 7).

The PCA confirms that the research questions are supported not only by individual visualizations but also by global multivariate structure in the data.

5.6 Interactive Story

The Interactive Story Lab serves as a narrative and exploratory component that integrates multiple visual analytics tools into a single interactive environment. Rather than answering a single research question, this section supports a deeper understanding of how different behavioural features interact across RQ1 and RQ2 in particular.

The animated story visualisation shows engagement behaviour under different follower–following rate scenarios, where the x-axis represents the scenario-adjusted follower–following rate and the y-axis represents average likes (see fig. 14). Through animation, helps us observe how engagement patterns evolve across account types and rank groups, making dynamic relationships more visible than in static plots.

Additional interactive views further enrich the analysis. The decision boundary explorer visualises predicted bot probabilities from a logistic regression model, which allows examine how combinations of features separate bots from humans (see fig. 13). The clustering explorer enables unsupervised grouping of accounts, highlighting whether natural clusters align with bot and human labels (see fig. 12). Finally, the detection sandbox uses parallel coordinate plots to compare real accounts with a user-defined hypothetical account, supporting intuitive multivariate reasoning (see fig. 11).

These interactive elements complement the static visualizations by enabling scenario-based exploration and reinforcing findings from RQ1 and RQ2.

5.7 AI-generated Graph

The graph was generated with the assistance of an AI tool, which was used to automatically produce a correlation heatmap from a subset of numerical features. The AI was provided with a structured prompt describing the dataset and the desired output. The prompt used was as follows:

> *"Given a dataset containing numerical features related to social media accounts, including follower–following rate, posting frequency, engagement metrics (likes, comments, reposts), URL usage, text length, and emotion indicators, generate a correlation heatmap that visualises the relationships between these variables. Use a diverging colour scale to distinguish positive and negative correlations."*

An AI-generated correlation heatmap is included to provide a high-level summary of relationships between numeric features in the dataset. This visualization is designed to complement the manually created plots by offering a global overview of how variables co-vary.

In the correlation heatmap, both axes represent numerical features, while colour intensity indicates the strength and direction of correlation (see fig. 9). Strong correlations between activity-related variables and engagement metrics support the findings from RQ2, while relationships involving follower–following rate and rank are consistent with observations from RQ1.

By presenting these relationships in a single visual, the AI-generated graph helps identify which features tend to move together and which provide complementary information. This supports exploratory reasoning and provides additional evidence that the behavioural differences highlighted throughout the dashboard are structurally embedded in the data.

6 Conclusion and Discussion

This project presented an interactive data visualization dashboard aimed at exploring behavioural differences between bot and human social media accounts. Using the SocialBot dataset, a wide range of visualisation techniques, including boxplots, histograms, scatter plots, principal component analysis, animated storytelling, and AI-assisted visual analytics, were applied to examine patterns related to popularity, activity, engagement, content style, and temporal behaviour.

The visual analysis revealed clear and consistent differences between bot and human accounts across multiple dimensions. Bots were found to exhibit more extreme follower-following ratios, lower user ranks, higher posting frequency, and lower engagement levels compared to human accounts. Content-based visualisations further showed that bot-generated posts tend to include URLs more frequently, display less linguistic variability, and contain fewer emotion-related tokens. Temporal analysis indicated that bots often follow more regular and predictable posting schedules, while human behaviour is characterised by greater variability. Together, these findings demonstrate that no single feature is sufficient to distinguish bots from humans instead meaningful separation emerges when multiple behavioural signals are considered jointly.

From a visualisation perspective, the project demonstrates the value of interactive and animated visual analytics in supporting exploratory data analysis and interpretation. The dashboard allows users to move beyond static summaries and actively engage with data through filtering, animation, and scenario-based exploration. In particular the animated story component helps communicate how engagement behaviour changes under different follower-following rate conditions, while the AI-generated correlation heatmap provides a high-level overview of the relationship among multiple numeric features.

Despite these strengths, the project also has limitations. The dataset represents aggregated account-level behaviour and does not capture post-level or longitudinal dynamics, which may obscure short-term temporal patterns or evolving behaviour. Additionally, some behavioural characteristics of bots and humans overlap, making perfect separation challenging. Future work could address these limitations by incorporating larger or more diverse datasets, analysis time-series data, or combining visual analytics with more advanced machine learning techniques to further improve interpretability and robustness.

In conclusion, this project shows that visualization driven analysis can play a critical role in understanding social media behaviour and supporting bot detection. By combining clear visual design, interactivity, animation, and AI-assisted analytics, the dashboard provides an effective and interpretable tool for exploring complex behavioural patterns. The approach highlights how data visualization can complement traditional analytical methods and contribute to more transparent and human-centered analysis of automated online activity.

7 Detailed Contribution

This section presents a detailed and transparent overview of individual contributions to both the dashboard development and the written report.

Table 1: Detailed Contribution of Group Members

Team Member	Project Contribution	Report Contribution
Fahim Shahriar	• Overall dashboard design and layout • Profile & Popularity tab • Activity & Engagement tab • Content & Timing tab • Interactive Story implementation	• Data Processing subsection • Visualization and Dashboard section • RQ1 – Profile and Popularity subsection • RQ2 – Activity and Engagement subsection • Interactive Story subsection
Adib Abzaal	• Activity & Engagement components • Principal Component Analysis (PCA) visualizations • AI-generated graph implementation	• Background and Motivation section • Project Objectives section • Data Source subsection • Data Description subsection • RQ3 – Content Style subsection • RQ4 – Temporal Patterns subsection • Principal Component Analysis (PCA) subsection • AI-generated Graph subsection • Conclusion and Discussion subsection

A Appendix: Additional Visualizations

A.1 Profile & Popularity (RQ1)

Figure 1: Follower–Following ratio comparison between Human and Bot accounts.

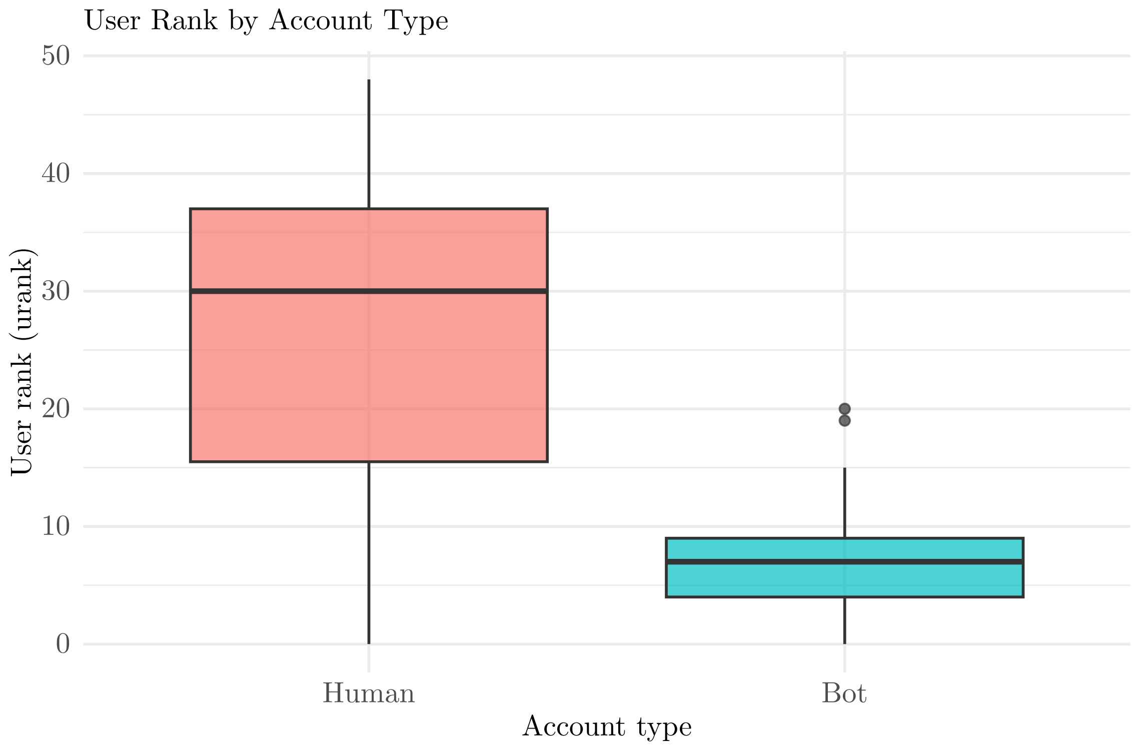

Figure 2: Distribution of user rank (urank) for Human and Bot accounts.

A.2 Activity & Engagement (RQ2)

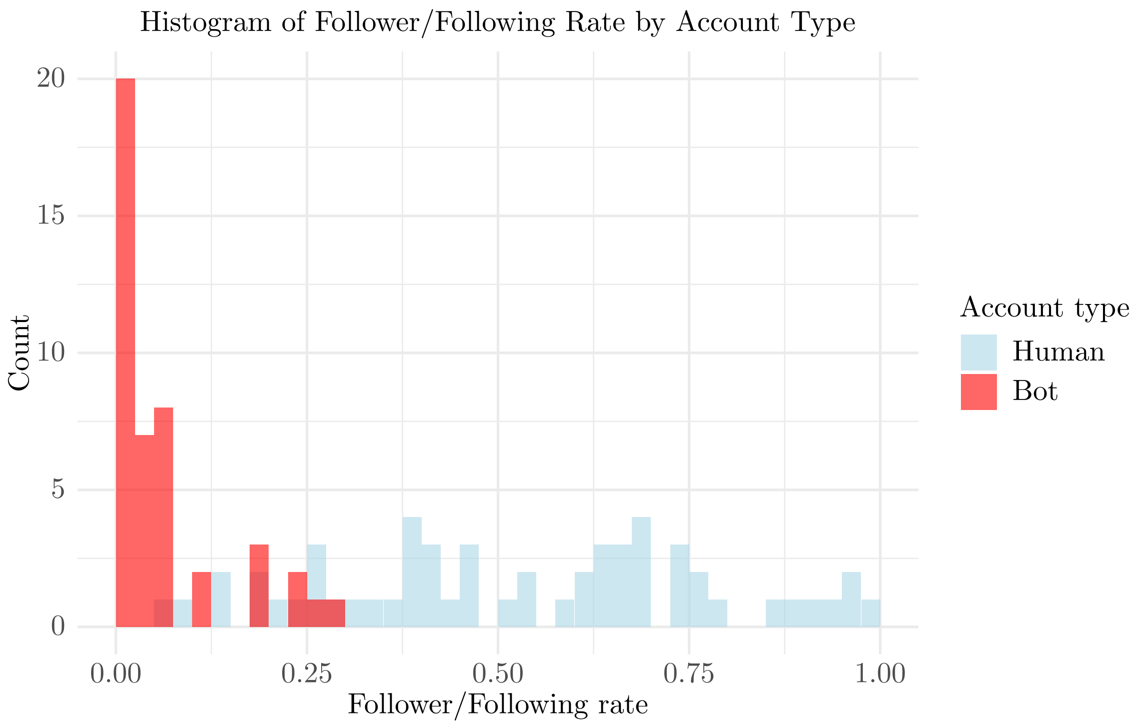

Figure 3: Histogram of follower–following rate by account type.

A.3 Content Style & Timing (RQ3–RQ4)

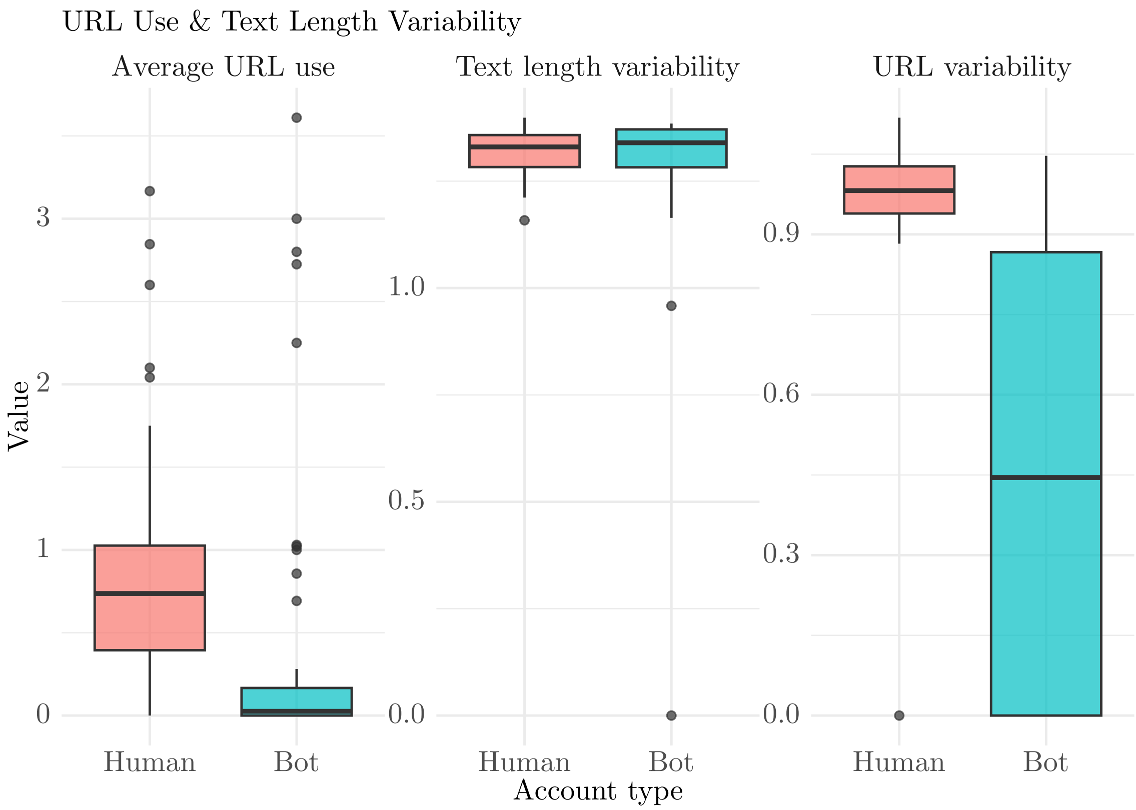

Figure 4: URL usage and text length variability across Human and Bot accounts.

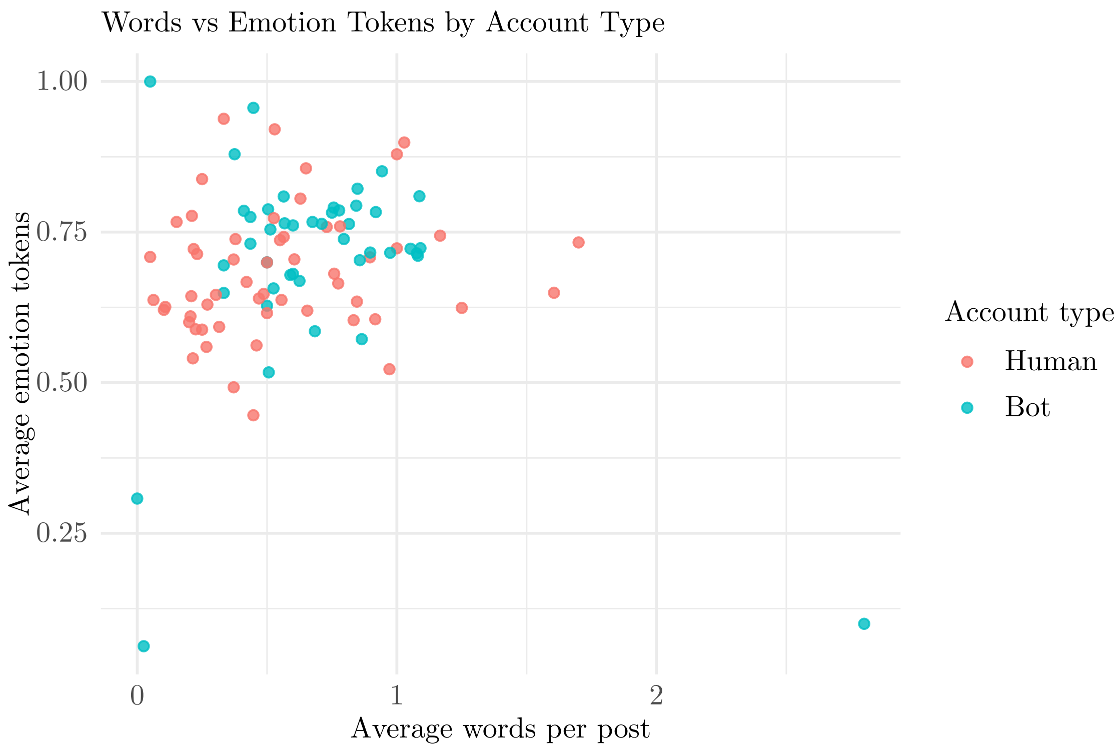

Figure 5: Relationship between average words per post and emotion tokens.

A.4 PCA Visualizations

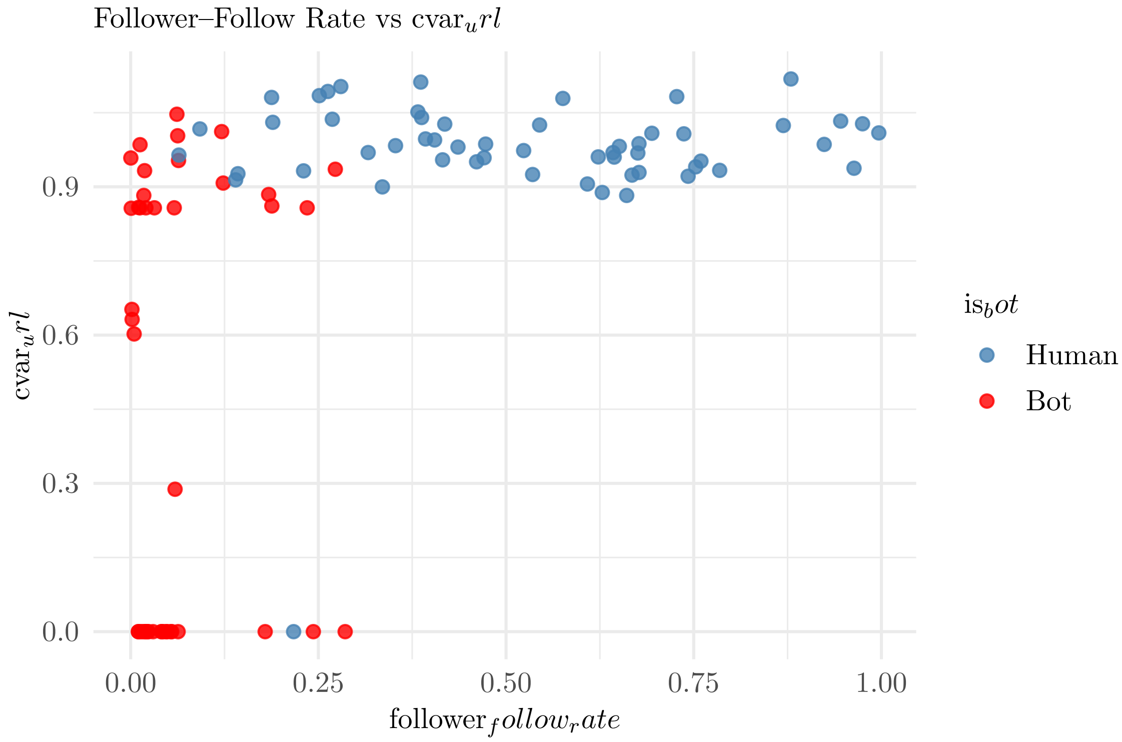

Figure 6: Pairwise scatter plot of follower–following rate vs URL variability (cvar_url).

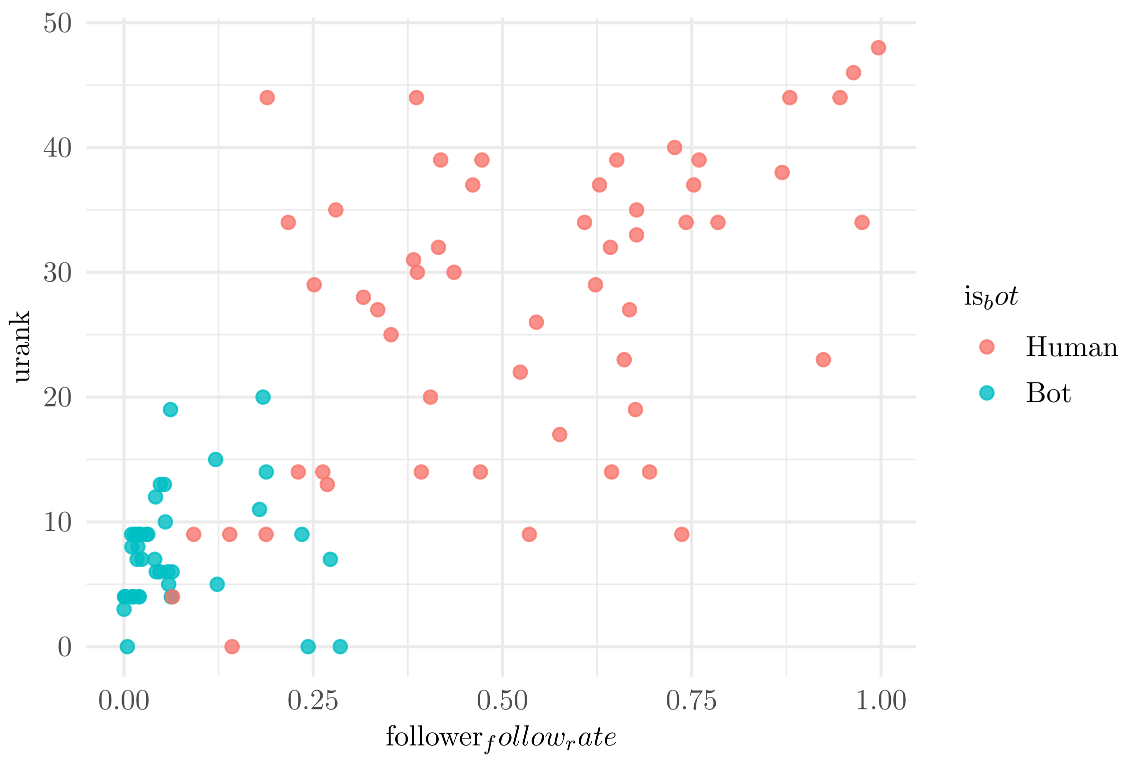

Figure 7: Pairwise scatter plot of follower–following rate vs user rank (urank).

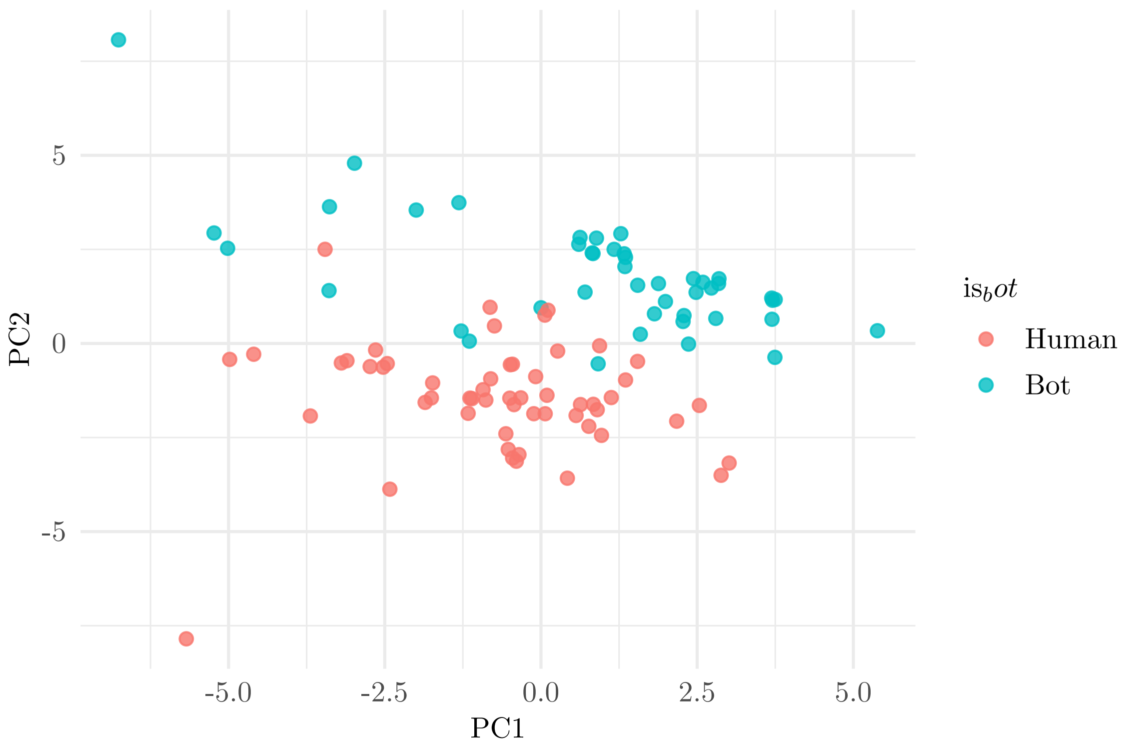

Figure 8: Principal Component Analysis (PC1 vs PC2) scatter plot coloured by account type.

A.5 AI-generated Correlation Heatmap

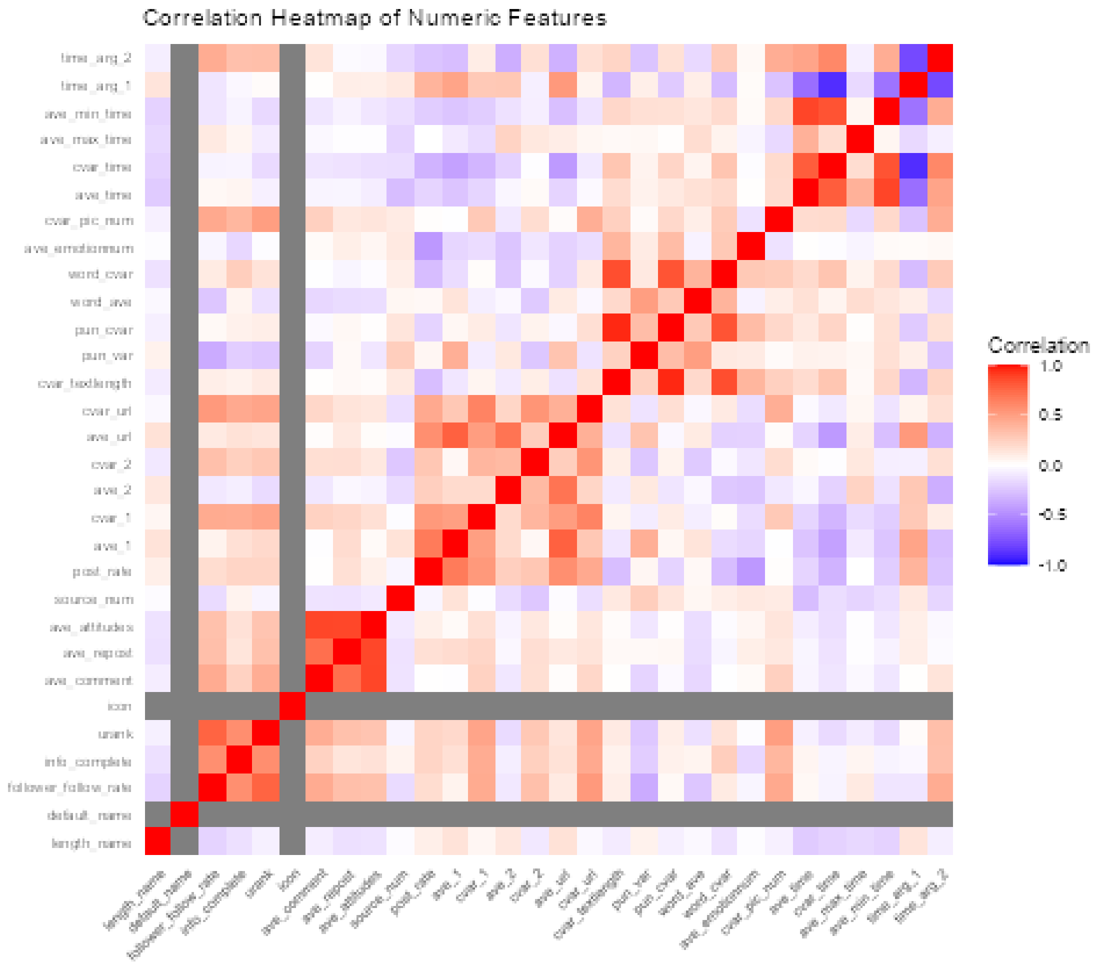

Figure 9: AI-generated correlation heatmap of numeric features, summarising global relationships between behavioural and content variables.

A.6 Interactive Story: Additional Screenshots

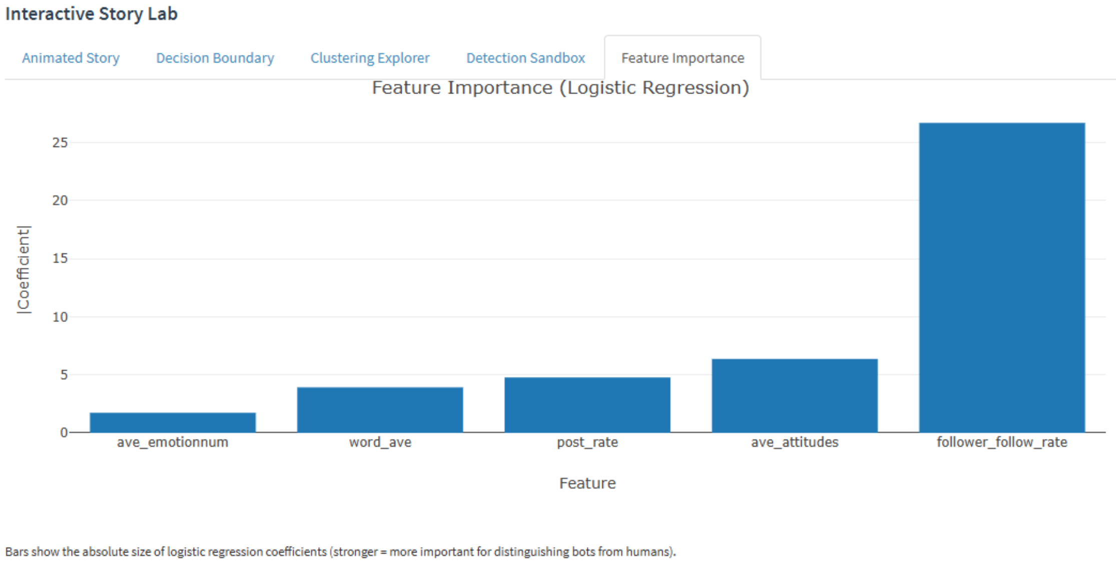

Figure 10: Feature importance derived from logistic regression. Bars represent the absolute magnitude of model coefficients, indicating the relative contribution of each feature to distinguishing bot and human accounts.

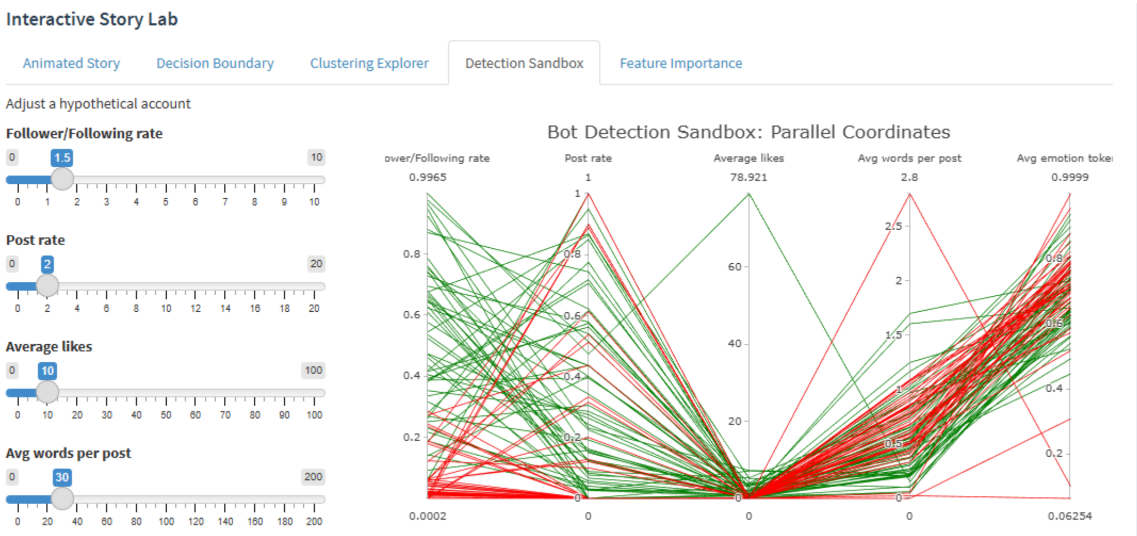

Figure 11: Bot Detection Sandbox using parallel coordinates. Each line represents an account, with colour indicating account type (Human or Bot). The highlighted scenario line shows a user-defined hypothetical account.

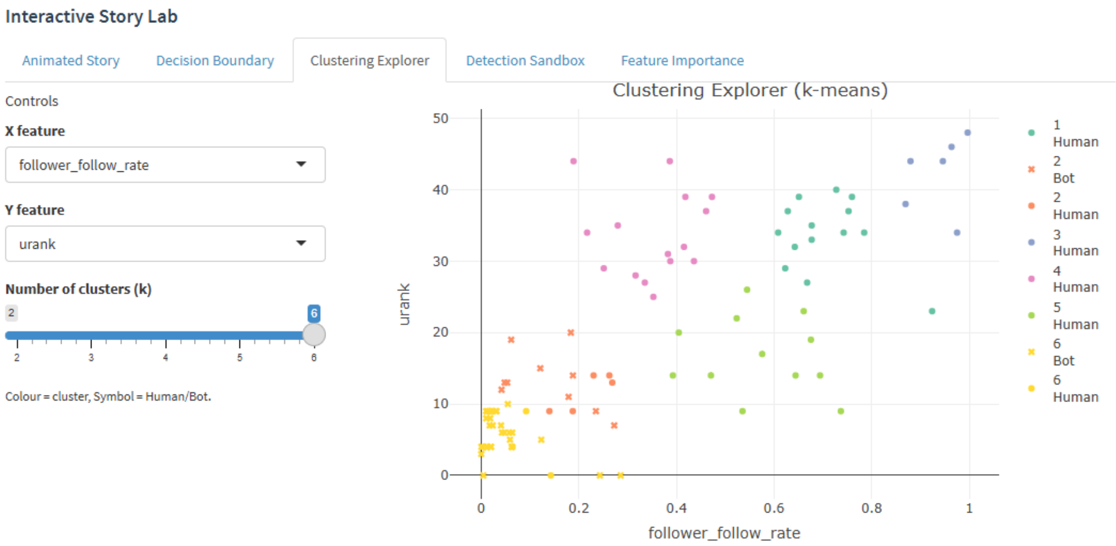

Figure 12: Clustering explorer based on k-means clustering. Points are coloured by cluster membership, while symbol shape indicates account type.

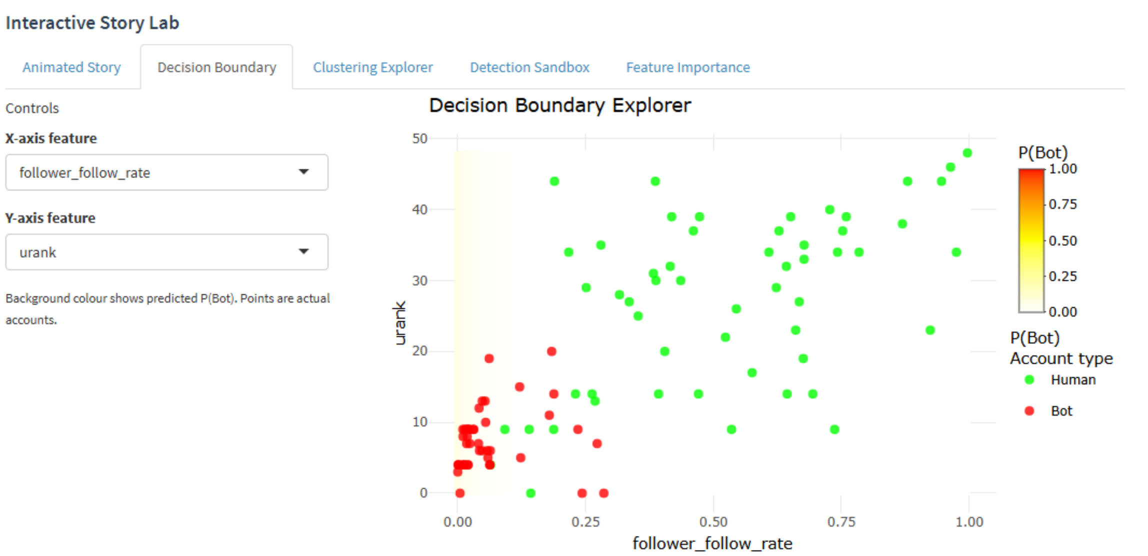

Figure 13: Decision boundary explorer based on a logistic regression model. Background colour represents predicted bot probability, while points show actual accounts.

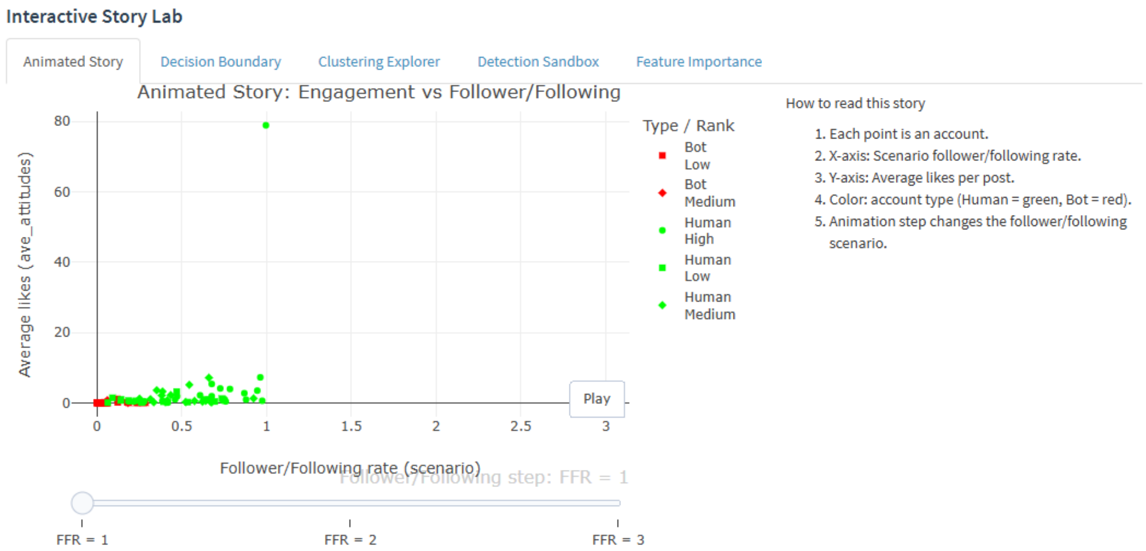

Figure 14: Animated story visualising engagement behaviour under different follower–following rate scenarios. Animation highlights how average likes vary across account types and rank groups.