

The background image shows the silhouette of the Hogwarts castle against a bright, circular light source, likely the moon, in a dark, cloudy sky. The castle's many spires and towers are visible, reflected in the dark water in the foreground.

Harry Potter NLP PROJECT

ALVIN HAMPTON

Project Overview: Analyzing Harry Potter Character Personalities

- To understand and categorize Harry Potter characters based on their dialogue
- Identify distinct personality archetypes within the wizarding world based on Big 5 personality traits and SLOAN system
- Dataset consists of the dialogue spoken from characters in all seven movies
- Any character who spoke any line in the movies appeared in the data (from 'Harry' to 'boy')

| | movie | chapter | character | dialog |
|------|---|----------------------|--------------------|---|
| 0 | Harry Potter and the Philosopher's Stone | Doorstep Delivery | Albus Dumbledore | I should have known that you would be here...P... |
| 1 | Harry Potter and the Philosopher's Stone | Doorstep Delivery | Minerva McGonagall | Good evening, Professor Dumbledore. Are the ru... |
| 2 | Harry Potter and the Philosopher's Stone | Doorstep Delivery | Albus Dumbledore | I'm afraid so, Professor. The good, and the bad. |
| 3 | Harry Potter and the Philosopher's Stone | Doorstep Delivery | Minerva McGonagall | And the boy? |
| 4 | Harry Potter and the Philosopher's Stone | Doorstep Delivery | Albus Dumbledore | Hagrid is bringing him. |
| ... | ... | ... | ... | ... |
| 7440 | Harry Potter and the Deathly Hallows Part 2 | Nineteen Years Later | Harry Potter | Then Slytherin House will have gained a wonder... |
| 7441 | Harry Potter and the Deathly Hallows Part 2 | Nineteen Years Later | Albus Potter | Really? |
| 7442 | Harry Potter and the Deathly Hallows Part 2 | Nineteen Years Later | Harry Potter | Really. |
| 7443 | Harry Potter and the Deathly Hallows Part 2 | Nineteen Years Later | Harry Potter | Ready? |
| 7444 | Harry Potter and the Deathly Hallows Part 2 | Nineteen Years Later | Albus Potter | Ready. |

7445 rows × 5 columns

First Steps: The model

- Use a pre-trained transformer model on the movie dialogue
- **Model type:** distilbert-base-uncased, fine-tuned
- **Trained on:** ~8,700 essays
- The model predicts big 5 personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism)
- Traits like Extraversion and Neuroticism had higher validation MSE, suggesting reduced predictive reliability.
- Cultural and linguistic biases in training data may influence predictions.

First Steps: The model

- Because the prediction model was trained on long-form texts, dialogue was aggregated into larger chunks prior to inference. This reduces noise associated with individual utterances and aligns the input distribution more closely with the model's training domain.
- Aggregated all dialogue for each character and then split these into 500-token chunks
- Calculated average scores for Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism across all chunks for each character .

Big 5 Personality Traits

- O.C.E.A.N
- The Big Five is the most scientifically robust personality framework. Unlike other methods, Big Five treats personality as continuous spectrums (you can score anywhere from low to high on each trait)
- Widely recognized model for understanding personality.
- Unlike MBTI or other methods, Big Five emerged from statistical analysis of language. Researchers studied how people describe themselves and others, then used factor analysis to identify five core dimensions that capture most personality variation.
- Empirically validated with thousands of studies across multiple countries confirm its reliability
- Correlates with job performance, relationship success as well as health outcomes
- remains relatively stable throughout most of one's lifetime and are influenced significantly by genes and the environment, with an estimated heritability of about 50%.
- <https://typemyvibe.ai/blog/big-five-guide>

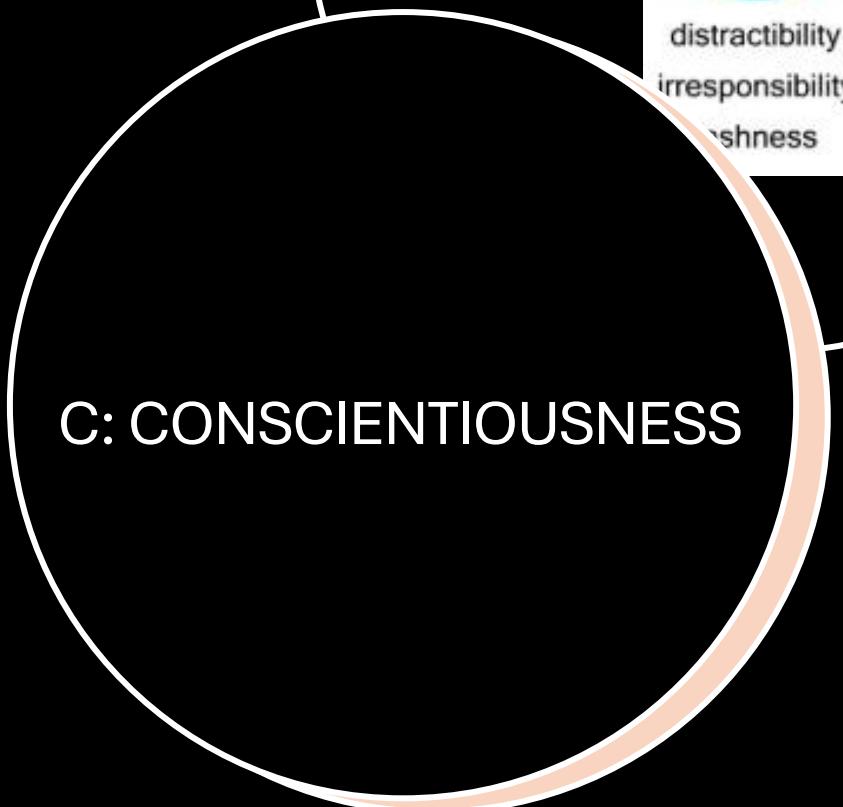
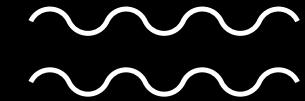
magical
thinking
eccentricity



inflexible
close-minded

O: OPENNESS

- This trait reflects a person's level of imagination, curiosity, intellectual depth, and preference for variety. Individuals scoring high enjoy novelty, abstract thinking, and art, while those scoring low prefer routine, tradition, and practicality.
- **HIGH**
- Curious and imaginative
- Enjoys abstract thinking
- Open to new experiences
- Creative
- **LOW**
- Practical
- Resistant to change
- Prefers routine
- conventional



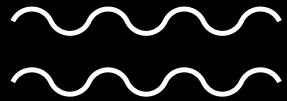
- This dimension measures a person's tendency toward self-discipline, organization, carefulness, and goal-directed behavior. A high score suggests a hardworking & reliable individual, whereas a low score points to someone who is more spontaneous, flexible, and perhaps prone to procrastination.
- **HIGH**
 - Disciplined
 - Detail-oriented
 - Responsible
- **LOW**
 - Spontaneous
 - Comfortable with disorder
 - Flexible



E: EXTRAVERSION

- This trait indicates how social, talkative, energetic, and assertive a person is, encompassing their level of engagement with the outside world. Highly extraverted people are energized by social interaction and crowds, while those scoring low (introverts) prefer solitude and reflection.
- **HIGH**
- Assertive
- Enthusiastic
- Energized by social interaction
- **LOW**
- Prefers solitude or small groups
- Reserved
- Calm

A: AGREEABLENESS



- This dimension gauges an individual's orientation toward compassion, cooperativeness, kindness, and trustworthiness in social interactions. High agreeableness is linked to being empathetic and helpful, while low agreeableness is associated with being more competitive, skeptical, and focused on self-interest.
- **HIGH**
 - Compassionate
 - Generous
 - Cooperative
- **LOW**
 - Direct
 - Competitive
 - Skeptical

depressivity

emotional
lability

shamefulness



fearlessness

shamelessness

N: NEUROTICISM

- This trait measures a person's tendency to experience negative emotions, such as anxiety, stress, worry, and emotional instability. A high score indicates high emotional reactivity and mood swings, while a low score suggests a calm, secure, and resilient temperament.
- High
- Sensitive to stress
- Self-conscious
- Anxious
- LOW
- Emotionally stable
- Confident
- Resilient to stress

Character Filtering

- After using model to generate Big Personality trait scores on everyone, characters were then filtered to remove less important characters before PCA and clustering
- It was important to remove less important characters **after** training the model because I wanted to establish a more ‘global average’. By training on everyone first (including extreme outliers like Vernon Dursley or Bellatrix Lestrange), the model learns the full range of character behavior in the script. This makes Harry’s high Neuroticism or Dumbledore’s high Openness stand out more accurately against the entire population.
- Importance was determined by amount of dialogue spoken throughout the series. If the chunk size (500 tokens per chunk) was less than 2, they were removed

Character Filtering: Impact on PCA

- Important to remove them before steps like PCA because if every minor character (like a Shopkeeper or witch) was included, their unique dialogue quirks might create "statistical noise."
- Characters with very little dialogue often have personality scores that are less reliable and more prone to random variation. If included, these noisy data points in PCA can distort the principal components.
- By removing them, it ensures that the PCA axes are defined by the most narratively significant differences.
- filtering beforehand ensures that PCA builds its components based on the most reliable and representative data available, leading to a more accurate and interpretable reduced-dimensional space for analysis.

Character Filtering: Impact on Clustering

- Background characters can act as "bridges." For example, a minor character might sit halfway between Cluster 1 and Cluster 4.
- If that "bridge" character is present, the clustering algorithm might merge those two distinct groups into one giant, messy cluster.
- By removing the "noise" characters, it allows a clear gap to form between the primary archetypes, resulting in much sharper, more distinct clusters that are easier to name and analyze.
- Since clustering often happens on the PCA-reduced data, having robust and clean principal components ensures that clusters are also more meaningful and less susceptible to the noise introduced by minor characters.

Scaling

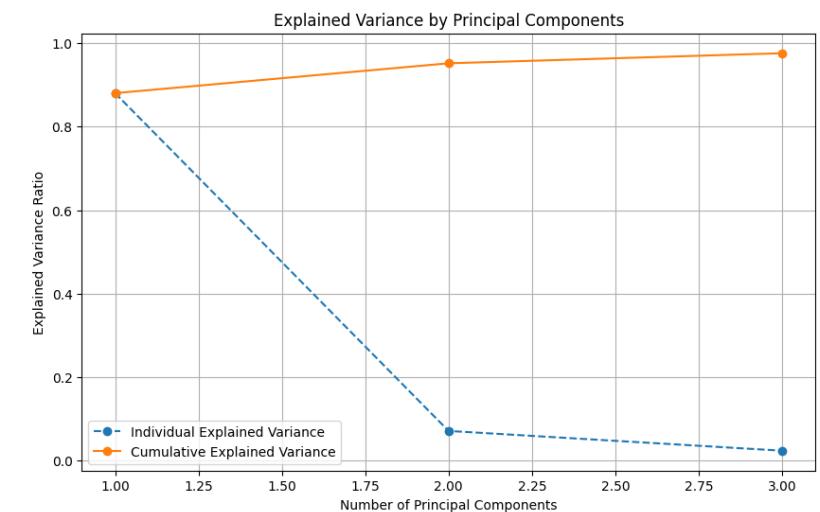
- Dataset has now been reduced to about 32 characters
- Each character has the average big 5 personality trait scores generated by the model
- Personality trait scores have different scales and variances, which could cause bias when using clustering algorithms.
- Standard Scaler applied to transform trait scores to have a mean of 0 and a standard deviation of 1
- This enables fair comparison between each character as well as meaningful clustering
- Scores represent deviation from the character population mean

Reasons to PCA before cluster

- One purpose of using Principal Component Analysis (PCA) in the workflow is to solve the problem of redundant, highly correlated features (Big 5 traits)
- If not, the cluster you get will be heavily defined by one dominant factor and fail to find groups on the nuance of others (openness or conscientiousness)
- Running PCA first, I can achieve 2 things: Decorrelation and Noise reduction
- **Decorrelation:** PCA transforms the original correlated features into a new set of uncorrelated, orthogonal features (PC1, PC2, etc.)
- **Noise Reduction:** can discard the later components (PC6, PC7, etc.) that only contain noise, focusing the clustering on the few components that capture the most meaningful variance.
- By clustering on the independent Principal Component scores, it ensures that **each dimension contributes fairly** to the grouping process, allowing the algorithm to find meaningful personality types based on different combinations of psychological factors.

Dimensionality Reduction (PCA):

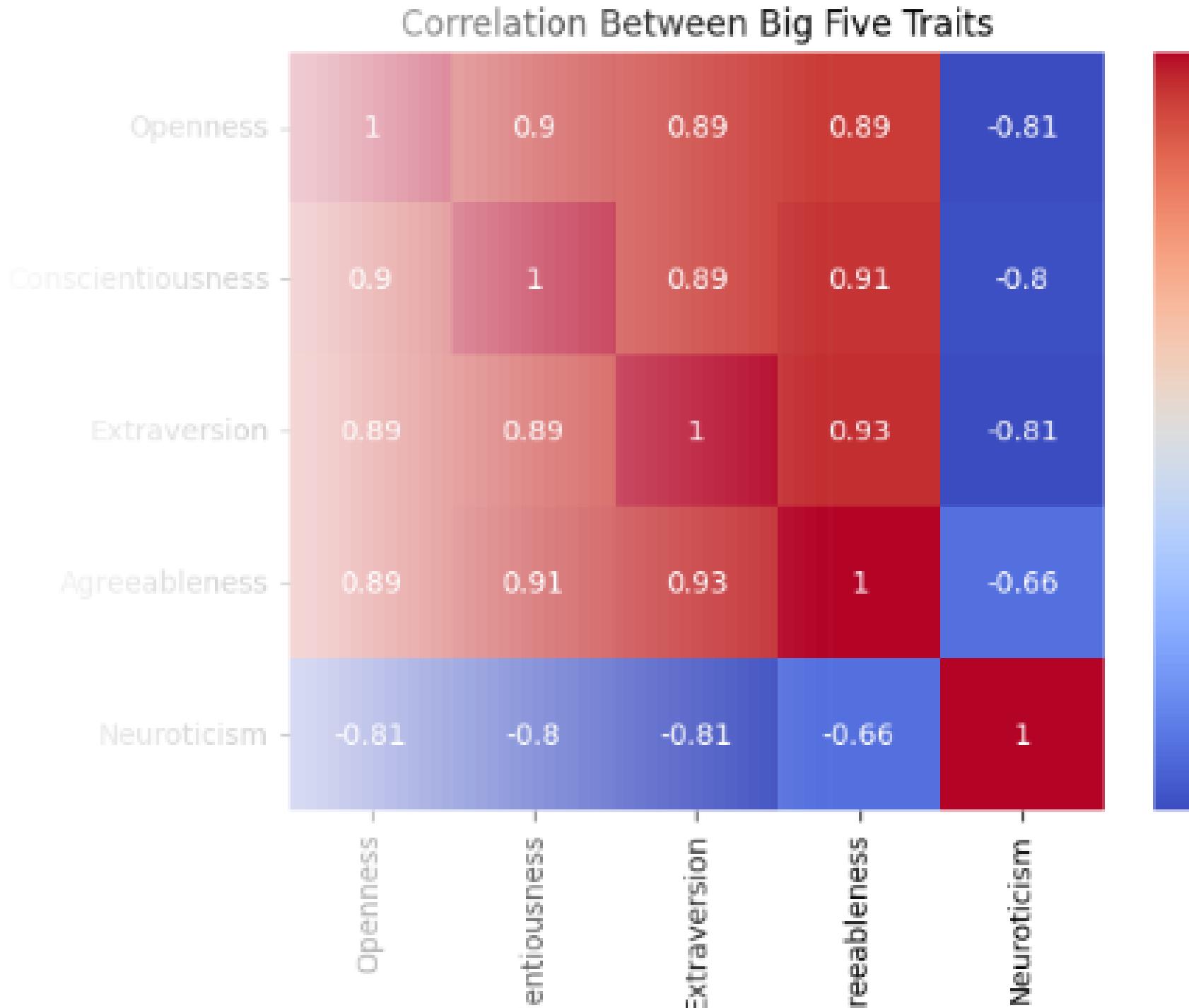
- Though only 5 columns, PCA used to simplify the data and focus on the most significant variance, making clustering more efficient and interpretable.
- Applied Principal Component Analysis (PCA) to the scaled trait scores, reducing the 5 dimensions to 3 principal components
- These 3 components collectively explained a high percentage of the total variance (e.g., ~97.57%)
- P_{c1}, P_{c2}, P_{c3} explaining ~88%, 7% and 2\$ respectively



—

Correlation

- Strong correlation between the scaled traits
- Essentially can be divided into 2 groups: Neurotic and Non-Neurotic traits
- This is confirmation of the high variance explanation of the first component



Loadings: Pc1

- Seems to be equally distributed with around more positive associated traits like Agreeableness, Conscientiousness, Openness and Extraversion
- Only negative loading in Pc1 is Neuroticism
- With 88% of variance being explained in this component, this can be looked at as the main divider of characters within their clusters
- Characters high in this component may be expected to have more authority and command while those with low scores may be less social and more reactionary

--- PC1 (88.01% of Variance Explained) ---

| PC1 | |
|--------------------------|-----------|
| Extraversion_scaled | 0.459385 |
| Conscientiousness_scaled | 0.457593 |
| Openness_scaled | 0.457161 |
| Agreeableness_scaled | 0.447439 |
| Neuroticism_scaled | -0.412765 |

| Character | pc1 |
|-----------------------|----------|
| 28 Sybill Trelawney | 4.533310 |
| 31 Voldemort | 3.129105 |
| 16 Lucius Malfoy | 2.696825 |
| 7 Dolores Umbridge | 2.514545 |
| 4 Bellatrix Lestrange | 2.509394 |
| 26 Severus Snape | 2.221285 |
| 1 Albus Dumbledore | 1.862426 |
| 29 Tom Riddle | 1.779541 |
| 6 Dobby | 1.601429 |
| 2 Argus Filch | 1.589355 |

Loadings: Pc2

- Extremely high Neuroticism as well as reasonably high agreeableness. Interesting, as these are often inversely related
- Extraversion and Openness are pretty much non contributors
- It measures the difference between characters who remain cool under pressure and those whose dialogue is emotionally charged, anxious, or defensive.

--- PC2 (7.15% of Variance Explained) ---

| PC2 | |
|--------------------------|----------|
| Neuroticism_scaled | 0.828767 |
| Agreeableness_scaled | 0.540950 |
| Conscientiousness_scaled | 0.113258 |
| Extraversion_scaled | 0.085493 |
| Openness_scaled | 0.019563 |

```
: df5[['Character', 'pc2']].sort_values(by='pc2', ascending = False).head(10)
```

| | Character | pc2 |
|----|--------------------|----------|
| 21 | Petunia Dursley | 1.232470 |
| 16 | Lucius Malfoy | 1.036930 |
| 10 | George Weasley | 0.858766 |
| 12 | Ginny Weasley | 0.852674 |
| 20 | Neville Longbottom | 0.776359 |
| 30 | Vernon Dursley | 0.636017 |
| 9 | Fred Weasley | 0.560027 |
| 6 | Dobby | 0.518893 |
| 7 | Dolores Umbridge | 0.318170 |
| 19 | Molly Weasley | 0.247239 |

Loadings: Pc3

- High positive Extraversion while also having noticeable negative scores for Openness and Conscientiousness
- Not much input for Neuroticism or Agreeableness
- May be the most complex group of characters who score high on this.
- Only explains about 2% of variation

| | Character | pc3 |
|----|--------------------|----------|
| 21 | Petunia Dursley | 1.232470 |
| 16 | Lucius Malfoy | 1.036930 |
| 10 | George Weasley | 0.858766 |
| 12 | Ginny Weasley | 0.852674 |
| 20 | Neville Longbottom | 0.776359 |
| 30 | Vernon Dursley | 0.636017 |
| 9 | Fred Weasley | 0.560027 |
| 6 | Dobby | 0.518893 |
| 7 | Dolores Umbridge | 0.318170 |
| 19 | Molly Weasley | 0.247239 |

| --- PC3 (2.41% of Variance Explained) --- | |
|---|-----------|
| | PC3 |
| Extraversion_scaled | 0.716077 |
| Openness_scaled | -0.497856 |
| Conscientiousness_scaled | -0.455558 |
| Agreeableness_scaled | 0.149490 |
| Neuroticism_scaled | -0.097435 |

Principal Component Loadings Takeaway

The Component Loadings reveal that the primary differentiator in the Harry Potter universe, regarding personalities, may not be the typical 'Good vs. Evil,' but rather Agency vs. Reactivity.

PC1 captures the character's ability to influence their environment, while PC2 seems to captures emotional resilience during that process.

Clustering: Hierarchical

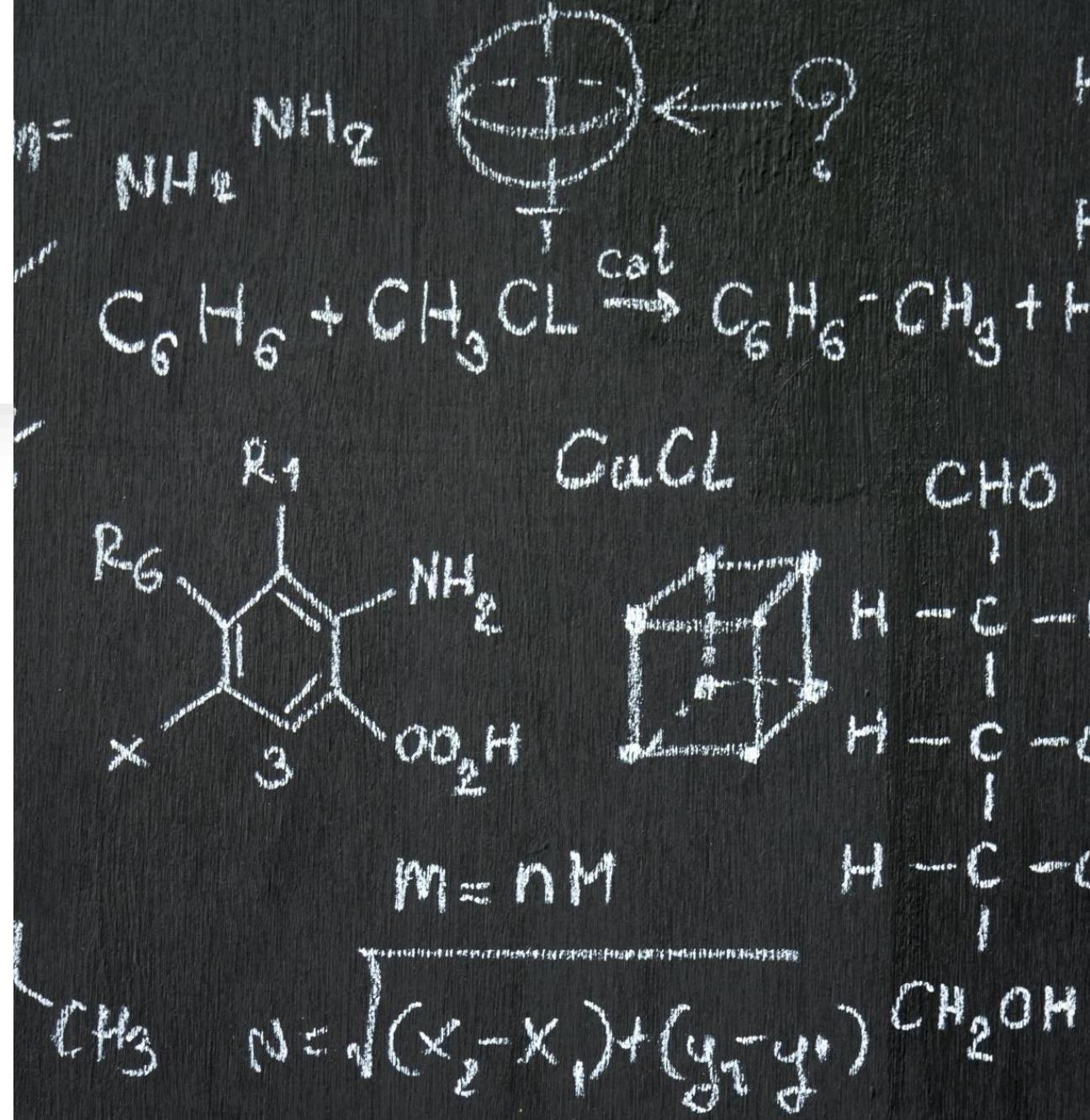
- I needed to see what would be the best clustering method
- For hierarchical, I opted with agglomerative clustering
- Using a combination of distance (Euclidean, Manhattan etc.), linkage (ward, complete, single etc.) and k ranging from 2 to 10, I plotted and calculate Silhouette, Calinski Harabasz, and Davies Bouldin scores see what gave me the best scores and made the most sense narratively

Clustering: K-Means

- K ranging from 2 to 10
- Calculated Silhouette Score and Inertia and plotted using elbow method to determine the optimal K
- Inertia is also plotted and calculated
- Inertia (WCSS or Within cluster sum of squares): This is the direct optimization objective of K-Means (it tries to minimize inertia). It measures cluster compactness.

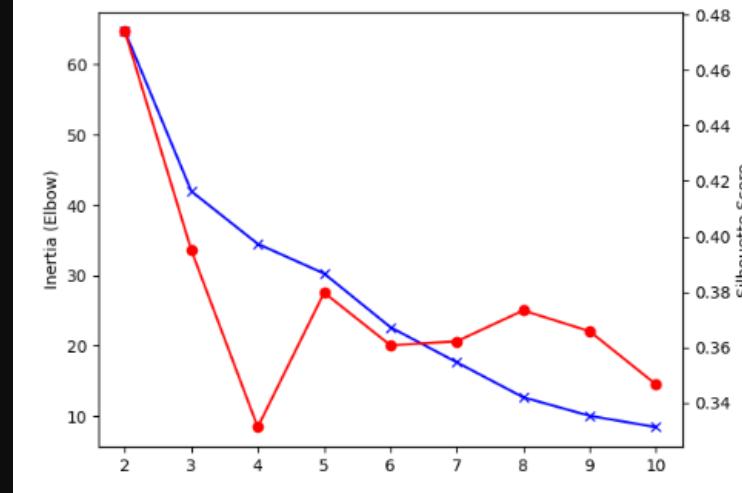
Scores

- **Silhouette Score:** tells you two things about each data point: how similar it is to other points in its **own** cluster (cohesion), and how different it is from points in *other* cluster s (separation). Range is from -1 to 1 with +1 being the best
- **Davies-Bouldin Score:** looks at how spread out the points are *within* each cluster (how 'messy' a group is), and how close the centers of different clusters are to each other (how 'overlapping' groups are). A lower Davies-Bouldin score indicates better clustering. It means your clusters are tight (not messy) and well-separated from each other.
- **Calinski-Harabasz Score:** This score looks at how spread out the clusters are *from each other* (how distinct they are) compared to how spread out the points are *within* each cluster (how tight they are). A higher Calinski-Harabasz score means better clustering. It suggests your clusters are distinct and internally dense. (**Note-** higher k tends to have higher C.H. score)



K-means vs Agglomerative Clustering

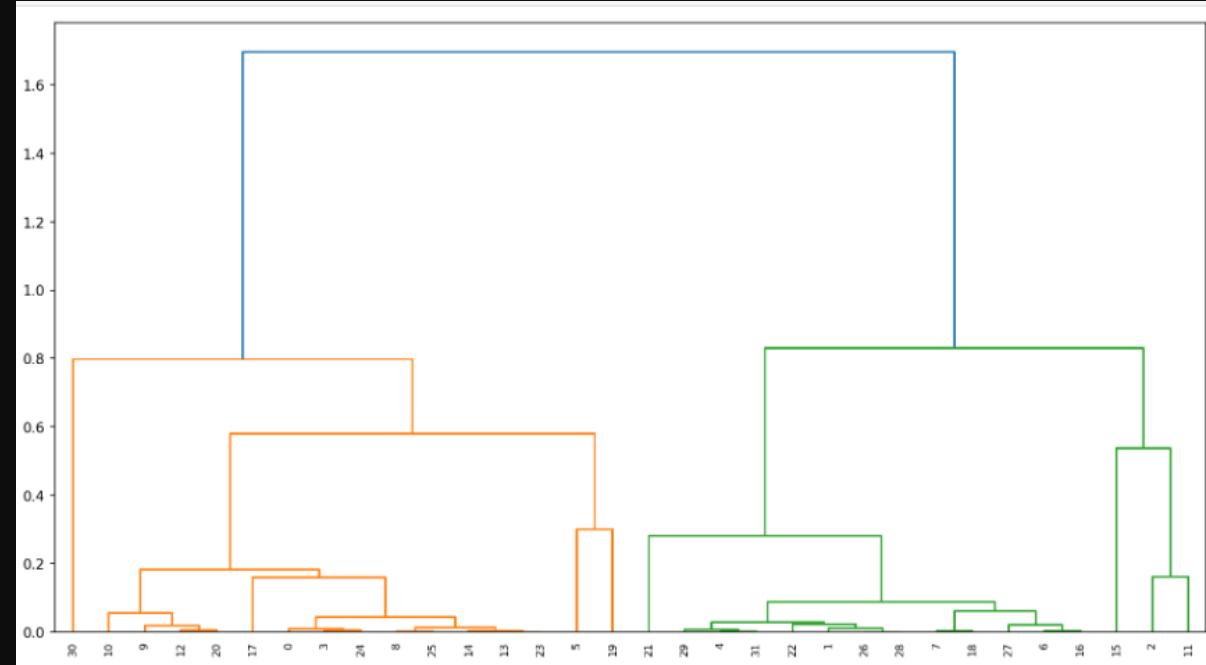
- With elbow method of K-means, inertia (blue) consistently gets lower as k increases
- Silhouette score (red) peaks at K=2 at around ~.50 (K-means)
- For Agglomerative Clustering, Silhouette score also peaks at 2 with ~.78.
- Agglomerative seems to be the best method in terms of clustering
- K = 3 may be the best option – Agglomerative
- Cosine is the best metric amongst all k-values. Cosine distance compares the angle between vectors, ignoring their magnitude. This not only makes sense, but it's perfect because it means that two characters with similar trait balance will cluster together even if one expresses traits more strongly than the other



| k | metric | linkage | silhouette | calinski_harabasz | davies_bouldin | n_clusters_found |
|----|--------|---------|------------|-------------------|----------------|------------------|
| 13 | 2 | cosine | complete | 0.782963 | 53.387525 | 0.658819 |
| 14 | 2 | cosine | average | 0.782963 | 53.387525 | 0.658819 |
| 29 | 3 | cosine | complete | 0.764019 | 33.152239 | 1.001574 |
| 30 | 3 | cosine | average | 0.764019 | 33.152239 | 1.001574 |
| 61 | 5 | cosine | complete | 0.721262 | 21.210002 | 0.854568 |
| 62 | 5 | cosine | average | 0.721262 | 21.210002 | 0.854568 |
| 13 | 6 | cosine | complete | 0.696136 | 17.152745 | 0.807657 |
| 14 | 6 | cosine | average | 0.696136 | 17.152745 | 0.807657 |
| 45 | 4 | cosine | complete | 0.695846 | 23.574601 | 0.886713 |
| 46 | 4 | cosine | average | 0.695846 | 23.574601 | 0.886713 |

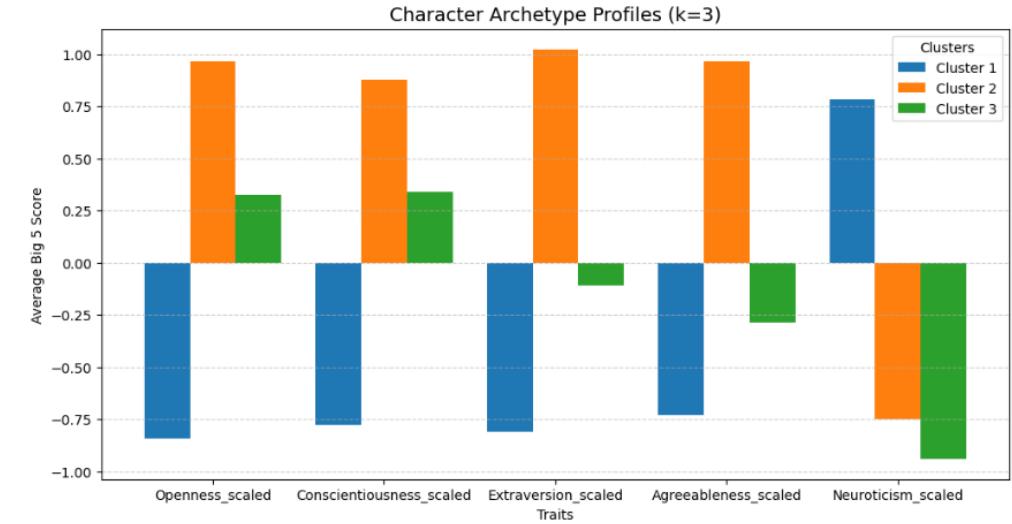
Clustering

- After removing generic speaker labels and retaining 32 named characters, the underlying personality structure became clear. K-Means clustering was evaluated as a baseline method; however, its assumptions of spherical, equally sized clusters were not well aligned with the structure of the PCA-transformed personality space.
- Hierarchical clustering using cosine distance produced stable and interpretable groupings across multiple cluster counts. In contrast, K-Means clustering showed no clear optimal partition, suggesting that the personality space is better characterized by hierarchical rather than centroid-based structure.



Cluster Analysis

- Cluster 1 and 2 seem to be the main clusters, with 3 having similar traits of both; small but cohesive and structurally consistent.
- This confirms what was seen when calculating metrics for clusters k = 2-10
- Cluster 1 and 2 are opposites.
- These characters within the groups should be on opposite ends of PC1 (~88% variance)
- Whatever traits makes up the PC1 should explain the division amongst clusters



| --- Average Personality Scores per Cluster --- | | | | |
|--|-----------------|--------------------------|---------------------|----|
| Final_Cluster | Openness_scaled | Conscientiousness_scaled | Extraversion_scaled | \\ |
| 1 | -0.845426 | -0.776051 | -0.811657 | |
| 2 | 0.964942 | 0.876434 | 1.024149 | |
| 3 | 0.327522 | 0.341058 | -0.109143 | |

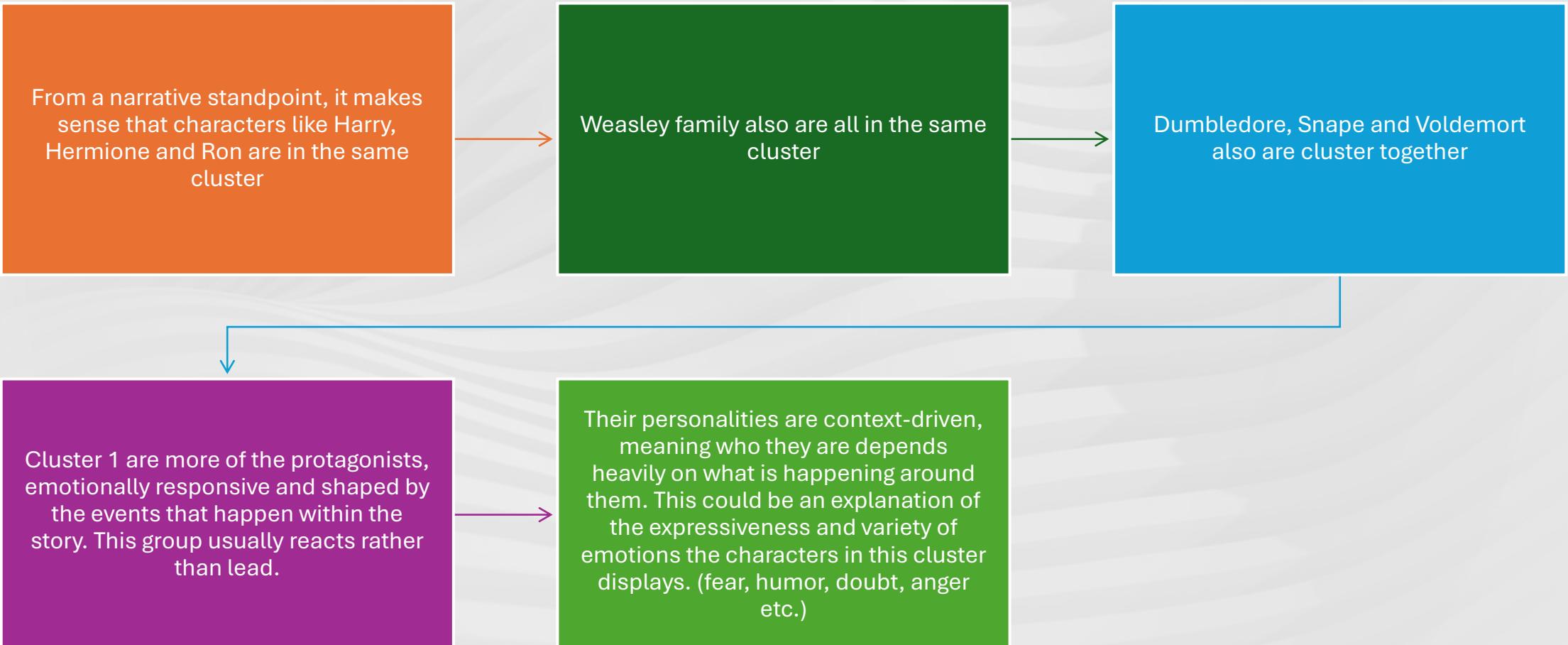
| Final_Cluster | Agreeableness_scaled | Neuroticism_scaled |
|---------------|----------------------|--------------------|
| 1 | -0.731203 | 0.784488 |
| 2 | 0.965960 | -0.748704 |
| 3 | -0.286079 | -0.939548 |

Clusters



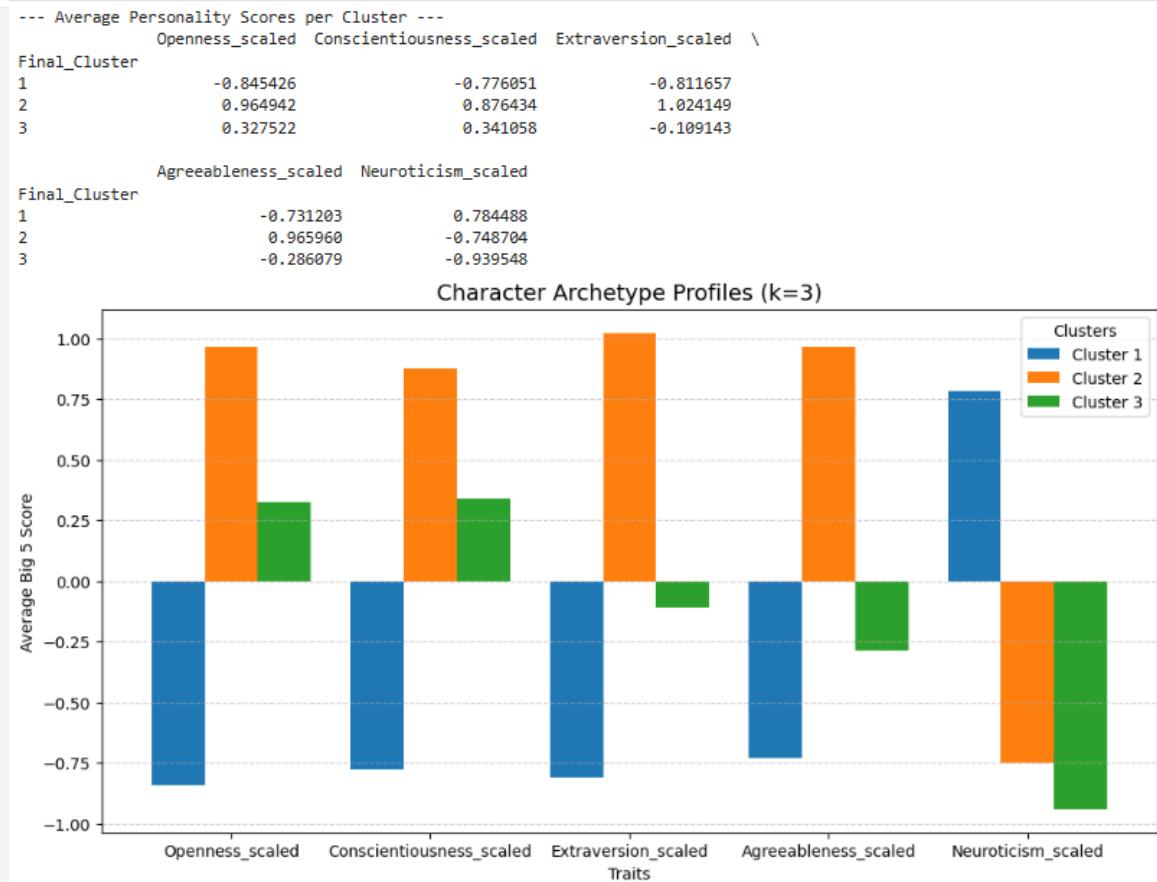
- **Cluster 1:** Alastor Moody, Arthur Weasley, Cornelius Fudge, Draco Malfoy, Fred Weasley, George Weasley, Ginny Weasley, Harry Potter, Hermione Granger, Luna Lovegood, Molly Weasley, Neville Longbottom, Ron Weasley, Rubeus Hagrid, Seamus Finnigan, Vernon Dursley
- **Cluster 2:** Albus Dumbledore, Bellatrix Lestrange, Dobby, Dolores Umbridge, Lucius Malfoy, Minerva McGonagall, Petunia Dursley, Remus Lupin, Severus Snape, Sirius Black, Sybill Trelawney, Tom Riddle, Voldemort
- **Cluster 3:** Argus Filch, Gilderoy Lockhart, Horace Slughorn

Clusters: Making Sense



Clusters: Making Sense

- Characters like Dumbledore, Dobby and McGonagall being clustered with Voldemort, Lucius and Bellatrix may seem odd at first glance
- If we reference the average Big 5 trait scores of cluster 2, we can see they have relatively high scores across the board except Neuroticism, (opposite of cluster 1)
- ‘Power players’, These are characters who have strong conviction in their beliefs
- High agency: they initiate, not react
- This cluster contains characters who define the direction of the story. Regardless of whether they are heroes or villains, they act based on deeply held beliefs or duties.
- Characters in this cluster seem to move the narrative forward instead of being moved by it. This is **THE** difference between the clusters



Clusters: Making Sense

Cluster 3 is a mix of cluster 1 and 2

Morally gray characters

Influence comes from position or reputation, not conviction

Confident, like cluster 2, but not as transformational or ideological.

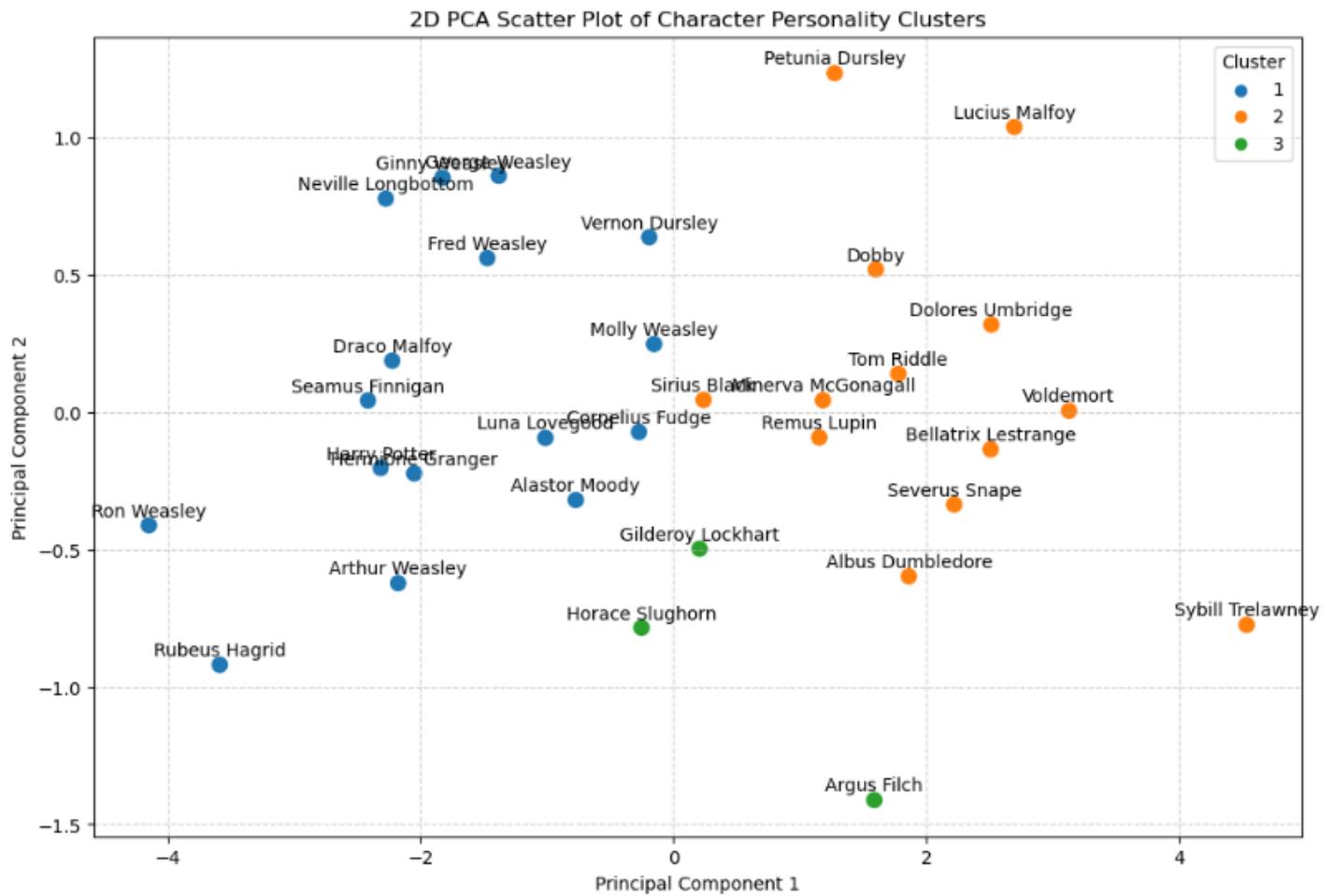
No leadership qualities, like Cluster 2 characters display, and not as emotionally reactive as Cluster 1 characters

These characters are more status or institution-oriented

This cluster does not seek to change the world nor respond emotionally to it; they seek to benefit from it.

Plot K=3

- with PC 1 and 2 explaining 95% variance, plotting should provide enough insight into character archetypes
- Cluster 1 and 2 are clearly on opposite ends of PC1



What's Happened

- Pre-trained transformer-based personality model used on Harry Potter movie dialogue to convert into big 5 personality scores
- Scores were standardized and characters were filtered based on amount of dialogue spoken
- PCA was applied on big 5 personality trait scores . This reduced noise, decorrelated traits and allowed easier visualization
- Clustering methods were determined by metrics like Inertia and Silhouette Score. Agglomerative Clustering methods was deemed the best method with optimal K =3

What's Next

1

Explore the SLOAN system

2

Create SLOAN codes for each character based on Big 5 personality trait scores

3

Explore SLOAN distribution

4

Cluster summaries and descriptions

5

Character analysis and deep dive

What are the SLOAN Dichotomies and How do they correspond to the Big 5 Traits?

The SLOAN system is an adaptation of the [Big 5 personality traits model](#). Now each of the Big 5 Traits is on a scale from High to Low, as a result, there are 10 different dichotomies. SLOAN gives each of these dichotomies a unique name and letter, making it easier to use.

Sloan & Big 5 Traits:

- High Extroversion = [Social](#) (S)
- Low Extroversion = [Reserved](#) (R)
- High Neuroticism = [Limbic](#) (L)
- Low Neuroticism = [Calm](#) (C)
- High Conscientiousness = [Organized](#) (O)
- Low Conscientiousness = [Unstructured](#) (U)
- High Agreeableness = [Accommodating](#) (A)
- Low Agreeableness = [Egocentric](#) (E)
- High Openness = [Inquisitive](#) (I)

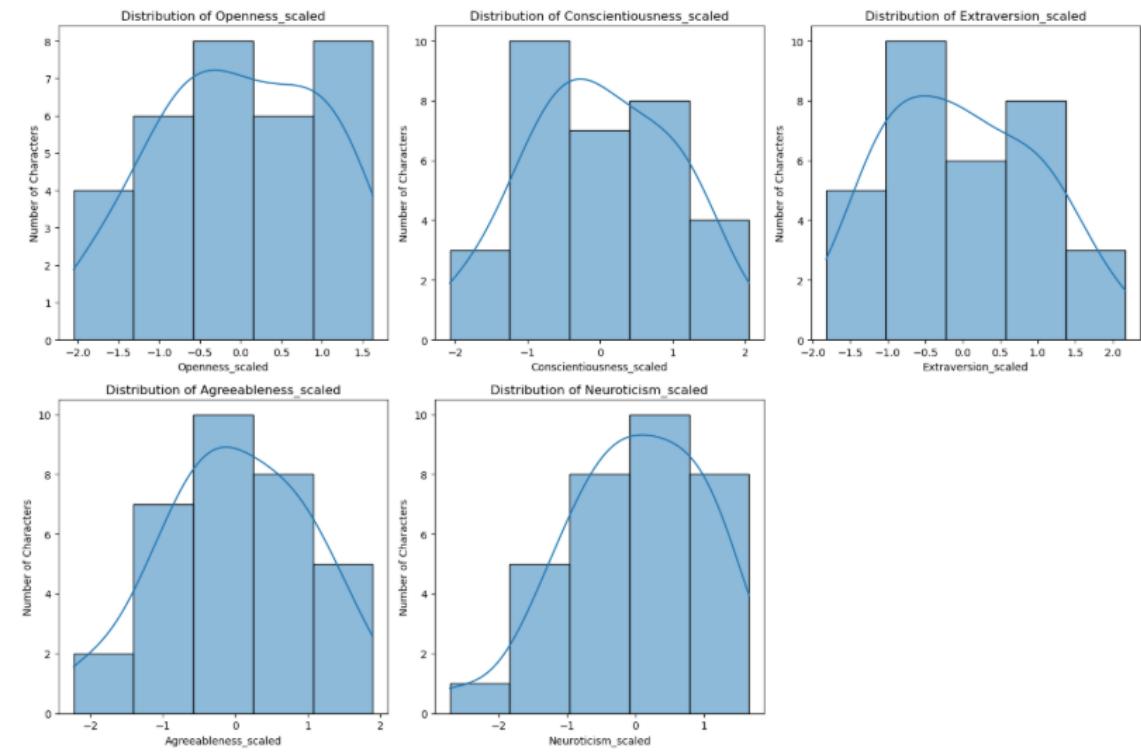
| | |
|--------------|---|
| <u>RCOAI</u> | Reserved (Introvert), Calm (Emotionally Stable), Ordered (High Conscientious), Accommodating (Agreeable), Inquisitive (High Openness) |
| <u>RCOAN</u> | Reserved (Introvert), Calm (Emotionally Stable), Ordered (High Conscientious), Accommodating (Agreeable), Non-curious (Low Openness) |
| <u>RLOAN</u> | Reserved (Introvert), Limbic (Neurotic), Ordered (High Conscientious), Accommodating (Agreeable), Non-curious (Low Openness) |
| <u>RLOAI</u> | Reserved (Introvert), Limbic (Neurotic), Ordered (High Conscientious), Accommodating (Agreeable), Inquisitive (High Openness) |
| <u>RLUAI</u> | Reserved (Introvert), Limbic (Neurotic), Unstructured (Low Conscientious), Accommodating (Agreeable), Inquisitive (High Openness) |
| <u>RLUAN</u> | Reserved (Introvert), Limbic (Neurotic), Unstructured (Low Conscientious), Accommodating (Agreeable), Non-curious (Low Openness) |
| <u>RCUAN</u> | Reserved (Introvert), Calm (Emotionally Stable), Unstructured (Low Conscientious), Accommodating (Agreeable), Non-curious (Low Openness) |
| <u>RCUAI</u> | Reserved (Introvert), Calm (Emotionally Stable), Unstructured (Low Conscientious), Accommodating (Agreeable), Inquisitive (High Openness) |
| <u>RCOEI</u> | Reserved (Introvert), Calm (Emotionally Stable), Ordered (High Conscientious), Egocentric (Disagreeable), Inquisitive (High Openness) |
| <u>RCOEN</u> | Reserved (Introvert), Calm (Emotionally Stable), Ordered (High Conscientious), Egocentric (Disagreeable), Non-curious (Low Openness) |
| <u>RLOEN</u> | Reserved (Introvert), Limbic (Neurotic), Ordered (High Conscientious), Egocentric (Disagreeable), Non-curious (Low Openness) |
| <u>RLOEI</u> | Reserved (Introvert), Limbic (Neurotic), Ordered (High Conscientious), Egocentric (Disagreeable), Inquisitive (High Openness) |

| | |
|--------------|---|
| <u>RLUEI</u> | Reserved (Introvert), Limbic (Neurotic), Unstructured (Low Conscientious), Egocentric (Disagreeable), Inquisitive (High Openness) |
| <u>RLUEN</u> | Reserved (Introvert), Limbic (Neurotic), Unstructured (Low Conscientious), Egocentric (Disagreeable), Non-curious (Low Openness) |
| <u>RCUEN</u> | Reserved (Introvert), Calm (Emotionally Stable), Unstructured (Low Conscientious), Egocentric (Disagreeable), Non-curious (Low Openness) |
| <u>RCUEI</u> | Reserved (Introvert), Calm (Emotionally Stable), Unstructured (Low Conscientious), Egocentric (Disagreeable), Inquisitive (High Openness) |
| <u>SCOAI</u> | Social (Extrovert), Calm (Emotionally Stable), Ordered (High Conscientious), Accommodating (Agreeable), Inquisitive (High Openness) |
| <u>SCOAN</u> | Social (Extrovert), Calm (Emotionally Stable), Ordered (High Conscientious), Accommodating (Agreeable), Non-curious (Low Openness) |
| <u>SLOAN</u> | Social (Extrovert), Limbic (Neurotic), Ordered (High Conscientious), Accommodating (Agreeable), Non-curious (Low Openness) |
| <u>SLOAI</u> | Social (Extrovert), Limbic (Neurotic), Ordered (High Conscientious), Accommodating (Agreeable), Inquisitive (High Openness) |
| <u>SLUAI</u> | Social (Extrovert), Limbic (Neurotic), Unstructured (Low Conscientious), Accommodating (Agreeable), Inquisitive (High Openness) |
| <u>SLUAN</u> | Social (Extrovert), Limbic (Neurotic), Unstructured (Low Conscientious), Accommodating (Agreeable), Non-curious (Low Openness) |
| <u>SCUAN</u> | Social (Extrovert), Calm (Emotionally Stable), Unstructured (Low Conscientious), Accommodating (Agreeable), Non-curious (Low Openness) |
| <u>SCUAI</u> | Social (Extrovert), Calm (Emotionally Stable), Unstructured (Low Conscientious), Accommodating (Agreeable), Inquisitive (High Openness) |
| <u>SCOEI</u> | Social (Extrovert), Calm (Emotionally Stable), Ordered (High Conscientious), Egocentric (Disagreeable), Inquisitive (High Openness) |

| | |
|--------------|---|
| <u>SLOEN</u> | Social (Extrovert), Limbic (Neurotic), Ordered (High Conscientious), Egocentric (Disagreeable), Non-curious (Low Openness) |
| <u>SLOEI</u> | Social (Extrovert), Limbic (Neurotic), Ordered (High Conscientious), Egocentric (Disagreeable), Inquisitive (High Openness) |
| <u>SLUEI</u> | Social (Extrovert), Limbic (Neurotic), Unstructured (Low Conscientious), Egocentric (Disagreeable), Inquisitive (High Openness) |
| <u>SLUEN</u> | Social (Extrovert), Limbic (Neurotic), Unstructured (Low Conscientious), Egocentric (Disagreeable), Non-curious (Low Openness) |
| <u>SCUEN</u> | Social (Extrovert), Calm (Emotionally Stable), Unstructured (Low Conscientious), Egocentric (Disagreeable), Non-curious (Low Openness) |
| <u>SCUEI</u> | Social (Extrovert), Calm (Emotionally Stable), Unstructured (Low Conscientious), Egocentric (Disagreeable), Inquisitive (High Openness) |
| <u>SCOEN</u> | Social (Extrovert), Calm (Emotionally Stable), Ordered (High Conscientious), Egocentric (Disagreeable), Non-curious (Low Openness) |

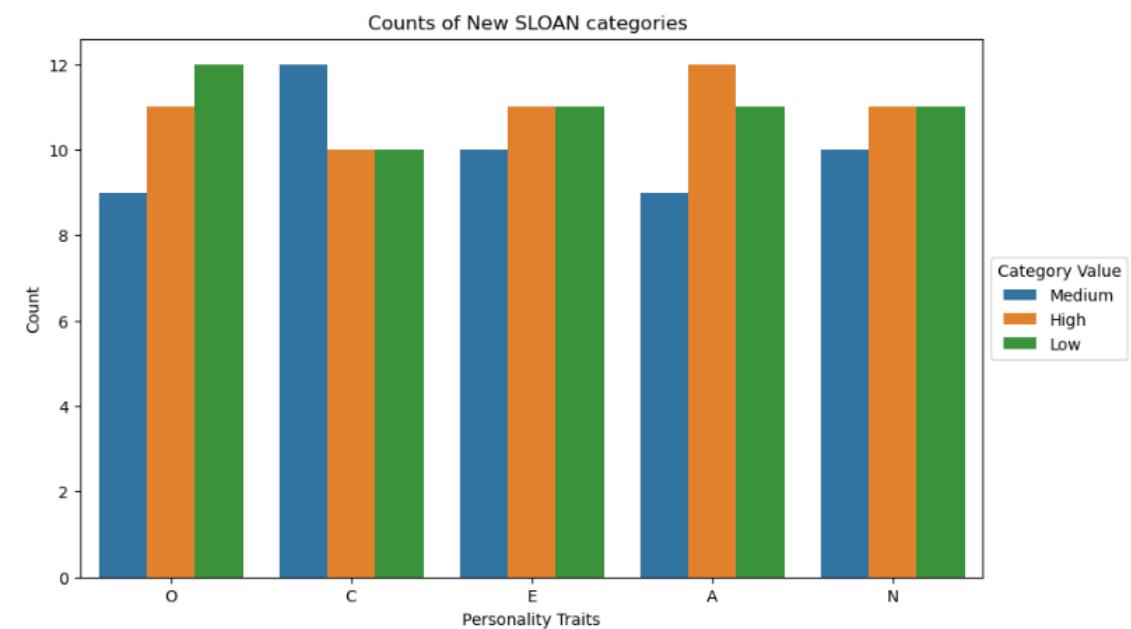
SLOAN CONVERSION

- Before conversion into SLOAN codes, personality scores must be categorized
- Big Five trait scores were discretized into low, medium, and high categories using fixed z-score thresholds (± 0.5).
- Low: < -0.5
- Medium: $-0.5 - +0.5$
- High: > 0.5
- Note: this is not inventing a signal but rather, only encoding information that already exists in the z-scores



SLOAN Conversion

- SLOAN codes are on a scale from Low to High.
- To improve interpretability while preserving signal strength, I extended SLOAN codes to distinguish strong trait deviations from moderate directional tendencies.
- **Uppercase letters denote high or low trait levels exceeding ± 0.5 standard deviations, while lowercase letters indicate directional tendencies within the medium range.**
- Standardized trait scores were discretized using fixed z-score thresholds (± 0.5).
- This resulted in a narrower ‘medium’ category centered around the population mean, ensuring that “high” and “low” labels reflect meaningful deviations rather than minor fluctuations.



SLOAN Conversion: Low Count

- Main characters and protagonists like Harry, Ron, Hermione and Draco have 4/5 traits categorized as Low
- Low trait characters, that so happen to include the Main Trio, can be seen as intentionally written that way
- These characters lack strong internal personality dominance but are emotionally reactive and relationally driven
- Their development happens through experience rather than ideology which may be the reason why they grow the most across the films

| | Character | High_Count | Medium_Count | Low_Count |
|----|---------------------|------------|--------------|-----------|
| 16 | Lucius Malfoy | 4 | 1 | 0 |
| 4 | Bellatrix Lestrange | 4 | 0 | 1 |
| 7 | Dolores Umbridge | 4 | 0 | 1 |
| 26 | Severus Snape | 4 | 0 | 1 |
| 28 | Sybill Trelawney | 4 | 0 | 1 |
| 29 | Tom Riddle | 4 | 0 | 1 |
| 31 | Voldemort | 4 | 0 | 1 |
| 6 | Dobby | 3 | 2 | 0 |
| 21 | Petunia Dursley | 3 | 2 | 0 |
| 1 | Albus Dumbledore | 3 | 1 | 1 |
| 22 | Remus Lupin | 3 | 1 | 1 |
| 18 | Minerva McGonagall | 2 | 3 | 0 |
| 2 | Argus Filch | 2 | 2 | 1 |
| 30 | Vernon Dursley | 1 | 4 | 0 |
| 9 | Fred Weasley | 1 | 2 | 2 |
| 10 | George Weasley | 1 | 2 | 2 |
| 12 | Ginny Weasley | 1 | 1 | 3 |
| 25 | Seamus Finnigan | 1 | 1 | 3 |
| 8 | Draco Malfoy | 1 | 0 | 4 |
| 13 | Harry Potter | 1 | 0 | 4 |
| 14 | Hermione Granger | 1 | 0 | 4 |
| 20 | Neville Longbottom | 1 | 0 | 4 |
| 23 | Ron Weasley | 1 | 0 | 4 |
| 24 | Rubeus Hagrid | 1 | 0 | 4 |
| 5 | Cornelius Fudge | 0 | 5 | 0 |
| 27 | Sirius Black | 0 | 5 | 0 |
| 0 | Alastor Moody | 0 | 4 | 1 |
| 11 | Gilderoy Lockhart | 0 | 4 | 1 |
| 19 | Molly Weasley | 0 | 4 | 1 |
| 17 | Luna Lovegood | 0 | 3 | 2 |
| 15 | Horace Slughorn | 0 | 2 | 3 |
| 3 | Arthur Weasley | 0 | 1 | 4 |

SLOAN Conversion: High Count

- Characters with 4/5 personality traits categorized as High are Lucius, Bellatrix, Snape and Voldemort
- Strong extreme personalities
- Their narrative roles as drivers of conflict or authority figures
- They are not subtle or situational but consistently act from well-defined internal drives.
- Tend to exert influence rather than adapt to circumstances.

| | Character | High_Count | Medium_Count | Low_Count |
|----|---------------------|------------|--------------|-----------|
| 16 | Lucius Malfoy | 4 | 1 | 0 |
| 4 | Bellatrix Lestrange | 4 | 0 | 1 |
| 7 | Dolores Umbridge | 4 | 0 | 1 |
| 26 | Severus Snape | 4 | 0 | 1 |
| 28 | Sybill Trelawney | 4 | 0 | 1 |
| 29 | Tom Riddle | 4 | 0 | 1 |
| 31 | Voldemort | 4 | 0 | 1 |
| 6 | Dobby | 3 | 2 | 0 |
| 21 | Petunia Dursley | 3 | 2 | 0 |
| 1 | Albus Dumbledore | 3 | 1 | 1 |
| 22 | Remus Lupin | 3 | 1 | 1 |
| 18 | Minerva McGonagall | 2 | 3 | 0 |
| 2 | Argus Filch | 2 | 2 | 1 |
| 30 | Vernon Dursley | 1 | 4 | 0 |
| 9 | Fred Weasley | 1 | 2 | 2 |
| 10 | George Weasley | 1 | 2 | 2 |
| 12 | Ginny Weasley | 1 | 1 | 3 |
| 25 | Seamus Finnigan | 1 | 1 | 3 |
| 8 | Draco Malfoy | 1 | 0 | 4 |
| 13 | Harry Potter | 1 | 0 | 4 |
| 14 | Hermione Granger | 1 | 0 | 4 |
| 20 | Neville Longbottom | 1 | 0 | 4 |
| 23 | Ron Weasley | 1 | 0 | 4 |
| 24 | Rubeus Hagrid | 1 | 0 | 4 |
| 5 | Cornelius Fudge | 0 | 5 | 0 |
| 27 | Sirius Black | 0 | 5 | 0 |
| 0 | Alastor Moody | 0 | 4 | 1 |
| 11 | Gilderoy Lockhart | 0 | 4 | 1 |
| 19 | Molly Weasley | 0 | 4 | 1 |
| 17 | Luna Lovegood | 0 | 3 | 2 |
| 15 | Horace Slughorn | 0 | 2 | 3 |
| 3 | Arthur Weasley | 0 | 1 | 4 |

SOFT SLOAN APPROACH

- Because the personality scores are standardized and model-generated, I wanted to avoid interpreting small numerical differences directly.
- Traits were discretized into low, medium, and high categories to capture meaningful deviations from the population mean.
- Trait counts were then used to characterize overall personality intensity without weighting any single trait. At the individual level, soft SLOAN encoding was used to preserve nuance
- High and Low traits are more defining, while medium represent more situational
- Medium bin: Scores in this range indicate that a trait is not strongly expressed relative to the population of characters. This also prevents characters from being falsely labeled as ‘high’ or ‘low’

== Cluster 1: Top 10 SLOAN codes ==

| Final_Cluster | Hybrid_SLOAN | count |
|---------------|--------------|-------|
| 0 | RLUEN | 6 |
| 1 | RLUeN | 1 |
| 2 | RLuEN | 1 |
| 3 | RLueN | 1 |
| 4 | RIUEN | 1 |
| 5 | RluEn | 1 |
| 6 | rLUeN | 1 |
| 7 | rLoai | 1 |
| 8 | rluEn | 1 |
| 9 | scuen | 1 |

== Cluster 2: Top 10 SLOAN codes ==

| Final_Cluster | Hybrid_SLOAN | count |
|---------------|--------------|-------|
| 11 | SCOAI | 6 |
| 12 | SCoAl | 2 |
| 13 | ScOAI | 1 |
| 14 | ScOAi | 1 |
| 15 | SIOAi | 1 |
| 16 | scoAl | 1 |
| 17 | sluai | 1 |

== Cluster 3: Top 10 SLOAN codes ==

| Final_Cluster | Hybrid_SLOAN | count |
|---------------|--------------|-------|
| 18 | rCOel | 1 |
| 19 | rCoen | 1 |
| 20 | sCUEi | 1 |

SLOAN Distribution: Cluster 1

- There are 32 characters remaining after filtering
- Cluster 1 has 16 (half) characters
- The most dominant SLOAN code is RLUEN [Reserved (Introvert), Limbic (Neurotic), Unstructured (Low Conscientious), Egocentric (Disagreeable), Non-curious (Low Openness)]
- 6 Characters have RLUEN (Ron, Harry, Hermione, Hagrid, Neville and Draco)
- 13/16 either have the exact pattern RLUEN or a variant of it which seems less of a coincidence and more of a sign of a consistent cluster

Sloan & Big 5 Traits:

- High Extroversion = Social (S)
- Low Extroversion = Reserved (R)
- High Neuroticism = Limbic (L)
- Low Neuroticism = Calm (C)
- High Conscientiousness = Organized (O)
- Low Conscientiousness = Unstructured (U)
- High Agreeableness = Accommodating (A)
- Low Agreeableness = Egocentric (E)
- High Openness = Inquisitive (I)
- Low Openness = Non-curious (N)

----- Cluster 1: SLOAN breakdown =====

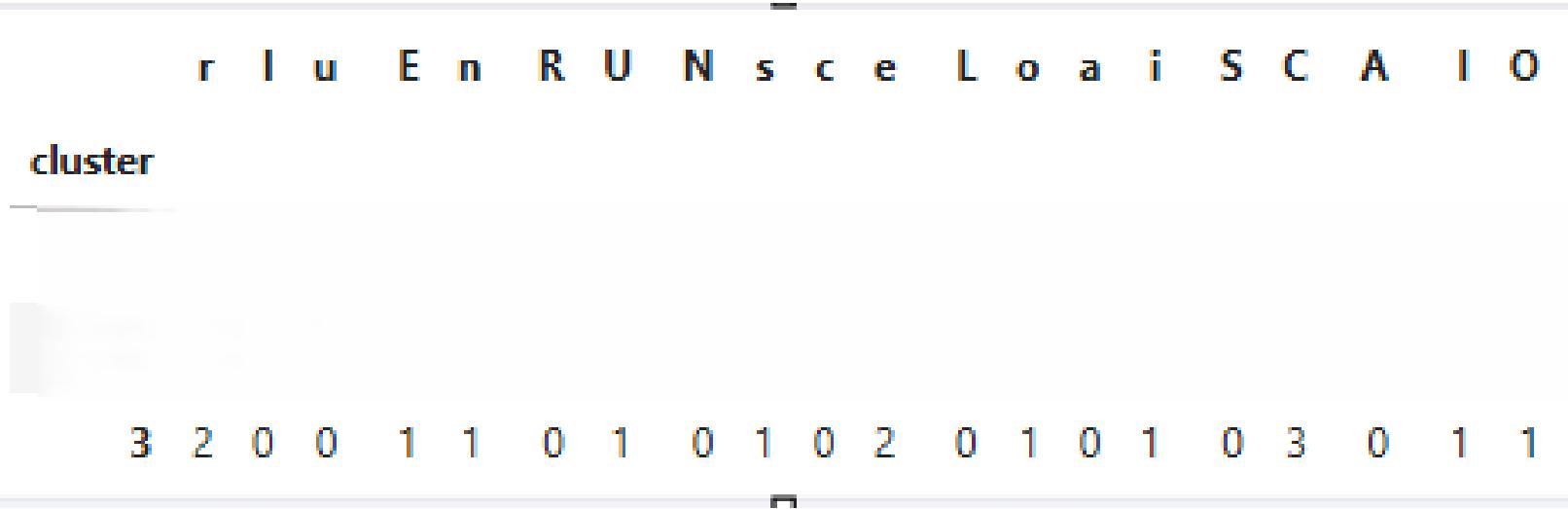
```
RLUEN : ['Draco Malfoy', 'Harry Potter', 'Hermione Granger', 'Neville Longbottom', 'Ron Weasley', 'Rubeus Hagrid']
RLUeN : ['Ginny Weasley']
RLuEN : ['Seamus Finnigan']
RLueN : ['Fred Weasley']
R1UEN : ['Arthur Weasley']
R1uEn : ['Luna Lovegood']
rLUeN : ['George Weasley']
rLoai : ['Vernon Dursley']
rluEn : ['Alastor Moody']
scuen : ['Cornelius Fudge']
sloaN : ['Molly Weasley']
```

SLOAN Distribution: Cluster 1



- R – Reserved: Inwardly focused, not necessarily selfish. Usually acts cautiously unless pushed. More reactor than initiator
- L - Limbic (Emotionally Reactive): The dominant trait in the cluster, these characters respond strongly to stress. Shaped by a wide range of emotions. Change behaviors based on situational context
- U – Unstructured: Situational behavior, inconsistent under pressure, reliance on others, improvisational decision-making
- E – Egocentric (low Agreeableness): Doesn't mean selfish in a moral sense but rather a strong personal perspective. Prioritization of loyalty. (Might explain Harry's rule-breaking, Draco's defensiveness and Hagrid's personal attachments)
- N – Non-Curious (low Openness): Preference for familiarity, discomfort with abstract ideologies and resistance towards changes on a large scale.

SLOAN Distribution: Cluster 3



- 3 characters make up this cluster: Argus Filch, Gilderoy Lockhart and Horace Slughorn
- Each with their own scores: rCOel, rCoEn, sCUEi respectively
- All characters have low Neuroticism (C for Calm): remain composed under pressure, rarely show emotional volatility, protect self image rather than emotionally react
- Low –Medium Agreeableness (E or e for Egocentric): A level of self-serving and recognition-seeking. Usually not acting based on morality
- Can be summarized as emotionally stable, status-oriented characters who seek comfort, recognition, or control within existing institution.. Their personalities are structured to maintain position, not to challenge systems or respond intensely to events.

SLOAN Distribution: Cluster 2

- 13 characters: Voldemort, Severus Snape, Bellatrix Lestrange, Dolores Umbridge, Tom Riddle, Sybill Trelawney, Remus Lupin, Dumbledore, Lucius Malfoy, Dobby, Sirius, Minerva McGonagall, Petunia Dursley
- Perhaps the most complex and conceptually important cluster from a narrative perspective
- 11 of 13 have a variant of the pattern SCOAI which is interesting since narratively these characters are on opposite sides
- SCOAI describes the capacity to act from conviction and sustain agency; it does not prescribe how that conviction is emotionally expressed.
- Not a villain cluster
- This can be seen as the most important cluster because it confirms what was suspected regarding the pattern the principal components recognized and what relationships the clusters discovered. It is **not about morality but psychological agency**

r | u E n R U N s c e L o a i S C A | O

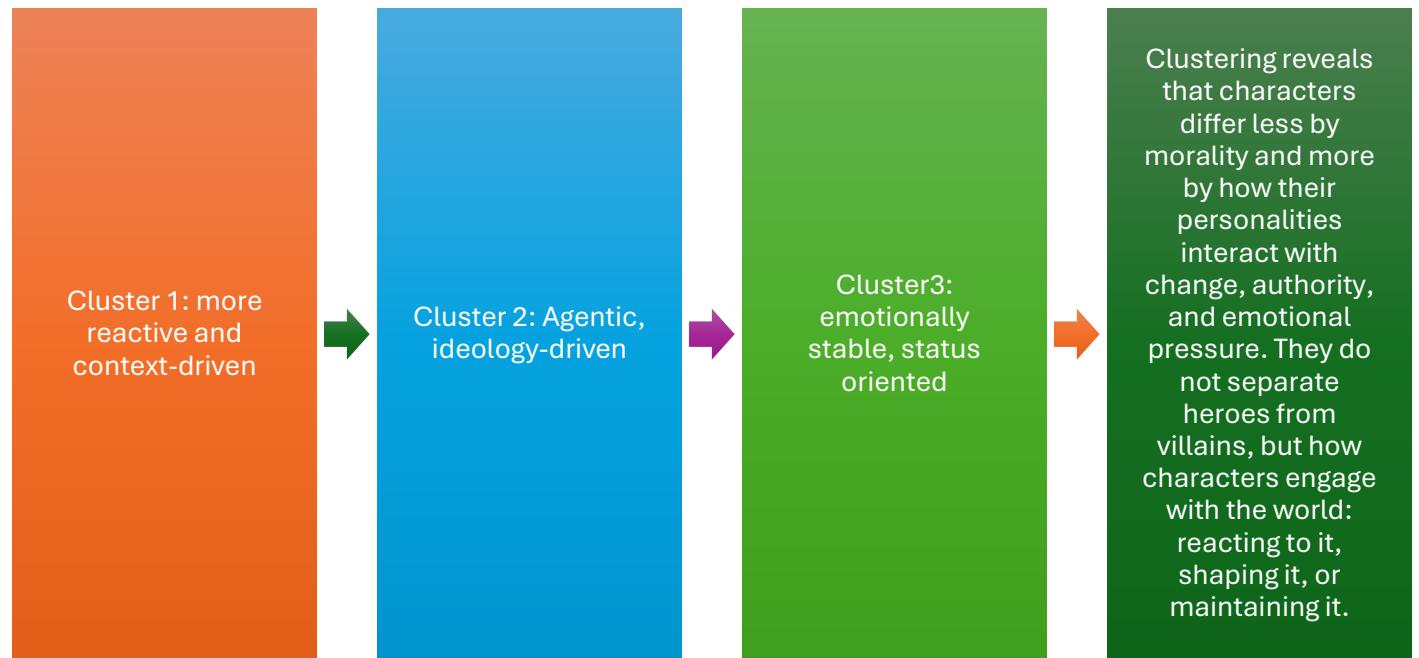
cluster

2 0 2 1 0 0 0 0 0 2 3 0 0 3 1 3 11 8 12 10 9

SLOAN Distribution: Cluster 2

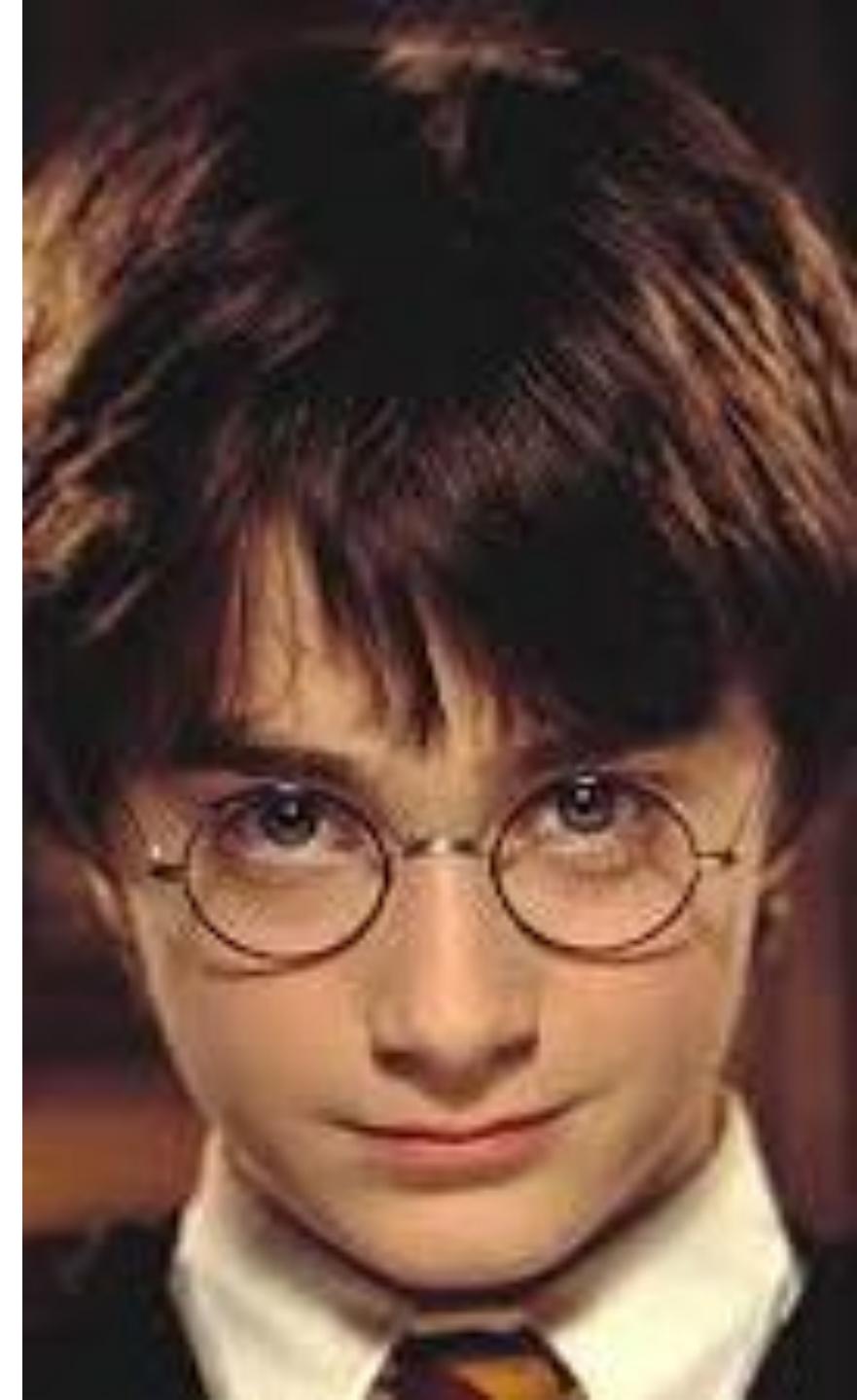
- A lot of strong traits with the strongest being A for Accommodating (high Agreeableness) and a close second being S for Social (high Extraversion)
- Noticeably high I (Inquisitiveness), O (Organized) and C (Calm) scores. The many uppercase variants shows may represent extreme internal clarity and certainty.
- ‘Power Players’ or authority figures who impose worldviews with low tolerance for ambiguity. Characters with rigid ideologies who operate from firm belief systems.
- Maintain consistent internal direction with the ability to influence others
- Cluster 2 is defined by high-agency characters whose actions are guided by internal conviction, ideology, or duty rather than emotional reaction or situational pressure. While their SLOAN profiles vary in expression, they share a capacity to initiate change, sustain long-term goals, and influence others.

Cluster Comparison and Review



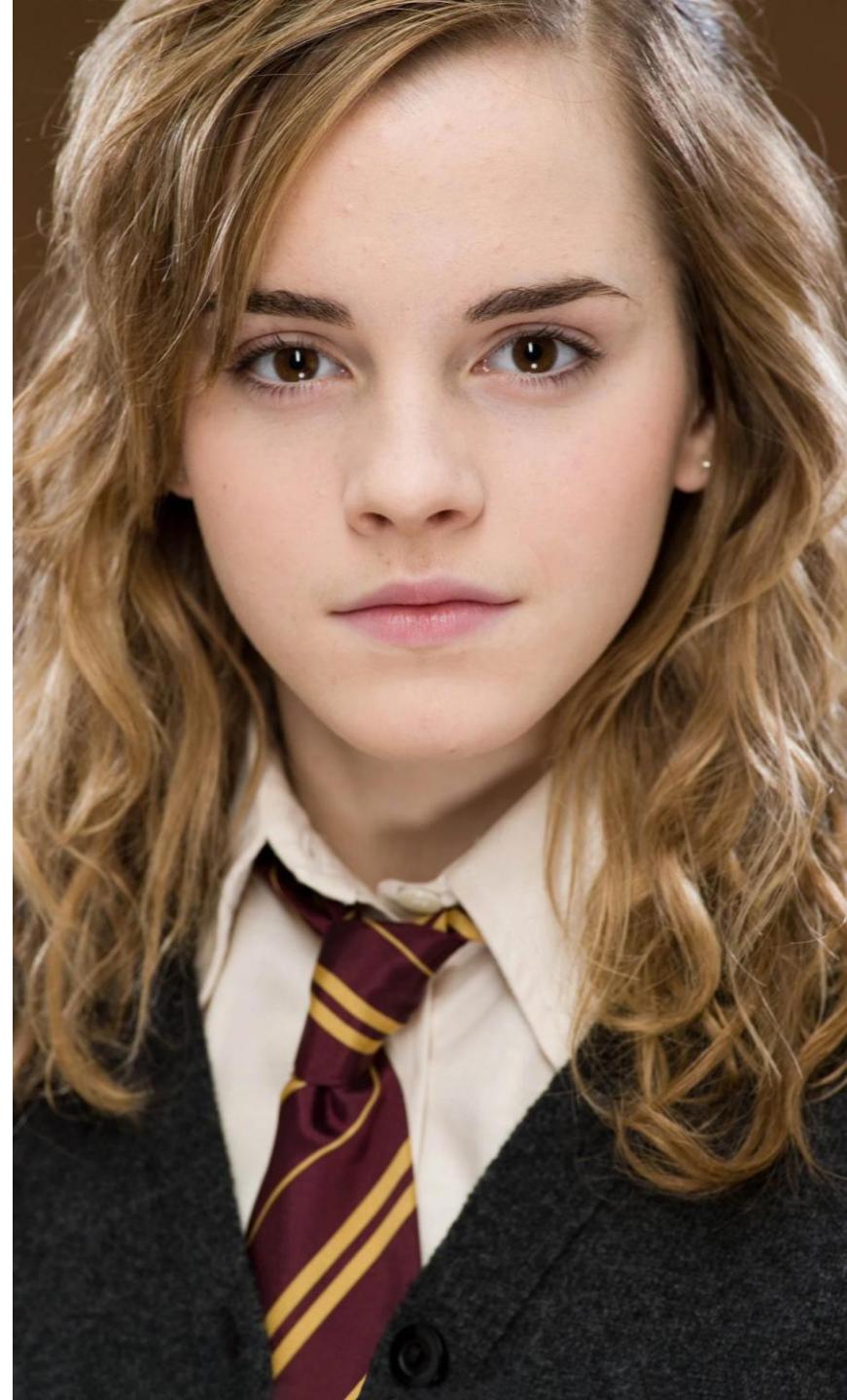
Character Descriptions: Harry Potter (RLUEN)

- Morally reactive
- expression is defined by moral immediacy, reacts emotionally to injustice and threats.
- Acts when emotionally compelled and not strategically motivated.
- Doesn't seek leadership, but steps forward because he feels the need to.
- His emotional reactivity fuels bravery rather than hesitation, but the underlying structure is still reactive, not agentic
- emotional sensitivity manifests as instinctive heroism.
- His high Reserved (R) score reflects a character who rarely initiates grand social acts; things happen **to** him.
- His Extreme Limbic (L) score is reflected in his dialogue. He is frequently angry, frustrated, or grieving and his words reflect the weight of his trauma.



Character Descriptions: Hermione Granger (RLUEN)

- Cognitive reactive
- Reacts to uncertainty by gathering information
- Emotional stress triggers rule-seeking and preparation
- appears organized, but her organization is reactive
- When order ceases, she becomes anxious rather than assertive, reinforcing the RLUEN profile.
- Her high Limbic (L) score captures her high-strung nature and her anxiety about getting expelled or failing a test
- Internally structured, used as a coping mechanism.
- She organizes to reduce uncertainty, not to control outcomes.
- though considered the "smart one", her movie dialogue is just as reactive to the plot as Harry and Ron's



Character Descriptions: Ron Weasley (RLUEN)

- Relational reactivity
- His high Limbic (L) score reflects how he wears his heart on his sleeve, whether it's fear of spiders, jealousy of Harry, or excitement. He is the most transparent character in the films.
- His high Egocentric (E) score isn't about being selfish but about his focus on his own immediate reality (hunger, discomfort, family status etc.)
- Emotional stress can manifest as jealousy, defensiveness, or withdrawal
- Emotional loyalty rather than principled
- Perceived social standing may shape his decisions
- emotional volatility can be seen as more interpersonal than moral or cognitive.
- emotional reactivity manifests as insecurity-driven behavior.



Character Descriptions: Draco Malfoy (RLUEN)

- Fear-driven emotional reactivity
- His behavior is shaped by expectation and pressure
- Emotional stress produces avoidance and hesitation
- May seem ideological but collapses under pressure because his personality lacks emotional regulation and internal stability
- His reactions expose anxiety rather than conviction
- emotional reactivity shows as defensive hostility



Character Descriptions: Remus Lupin (SCoAI)

- uses his intellect to connect with others (specifically Harry)
- Sophisticated dialogue and the only other character with same SLOAN code as Dumbledore
- May possess the same intellect as other highly viewed characters like but his moderate organized (o) score makes him feel more approachable and flexible than other authority figures.
- His high Calm (C) and Accommodating (A) scores reflect his role as the empathetic mentor who remains composed despite events or discussions of his past



Character Descriptions: Alastor Moody (rluEn)

- profile is defined by a single extreme: Egocentric (E), which doesn't mean "selfish", but uncompromising
- Pragmatist by nature, who's objective overrides manners occasionally
- Moderate scores in everything else reflect a seasoned veteran
- Disciplined and functional control over his speech unlike other characters with more extreme scores resulting in emotional outbursts



Character Descriptions: Gilderoy Lockhart (rCoen)

- Moderate Reserved, Extreme Calm, Moderate Organized, Moderate Egocentric, Moderate Non-curious
- He speaks confidently but avoids responsibility
- Planning is superficial and short-lived
- He does not experience anxiety when exposed
- Failure does not destabilize his self-image. He rationalizes instead of reflect
- His only extreme is Calm (C). In the movies, this represents his delusion. He is never rattled or anxious because he is entirely encapsulated by his own vanity



Character Descriptions: Minerva McGonagall (scoAI)

- Moderate Social, Moderate Calm, Moderate Organized, high Accommodating, high Inquisitive
- Her SLOAN code highlight her role as an academic authority who is deeply committed to the social order of Hogwarts
- Due to her Social, Calm, and Organized scores being moderate (sco), she feels more "human" and grounded than her contemporaries like Dumbledore or Snape
- She enforces rules even when emotionally costly. Maintains authority without emotional escalation
- For her, structure is not situational, it is identity-based. A notable difference to someone like Hermione who organizes to feel safe; McGonagall organizes because it is necessary.
- Stress does not destabilize her language or behavior



Character Descriptions: Lucius Malfoy (ScOAI)

- Extreme Social, Moderate Calm, Extreme Organized, Extreme Accommodating, Extreme Inquisitive
- structured behavior can be attributed to external systems, not internal conviction
- His discipline is strongest when status protects him. Composure relies on external backing
- Unlike Voldemort or Snape, Lucius has a moderate calm (c). This matches his arc in the films, where he starts as a cold, unshakeable figure. As he starts to lose favor with Voldemort, his composure begins to falter. The lowercase "c" captures that underlying desperation and fear that eventually peeks through his once polished exterior
- ***NOTE*** A difference between him and McGonagall is she enforces structure regardless of power while Lucius relies on structure to preserve power. McGonagall is the system; Lucius hides behind it.



Character Descriptions: Bellatrix Lestrange (SCOAI)

- high Social, high Calm, high Organized, high Accommodating, high Inquisitive.
- It may seem counter-intuitive that a "mad" character has an Extreme Calm (C) and Organized (O) score, but in the movies, she is never truly rattled. She is fearless. Her dialogue is not reactive or anxious; it is predatory.
- She does not hesitate, doubt, or second-guess; her actions flow directly from conviction.
- Extreme Social (S) impact is noticeable when present on the screen. Her Extreme Accommodating (A) trait is linked to her devotion to Voldemort; she is "accommodating" to his will above all else
- *NOTE* a difference between her and Snape. Where Snape controls emotion to preserve secrecy, Bellatrix unleashes emotion to demonstrate loyalty. Both act from belief, but Bellatrix's personality externalizes that belief through visible passion and aggression.



Character Descriptions: Sirius Black (sluai)

- While he sits in the Cluster 2 , the cluster mostly extreme scores, his movie dialogue is unique because it doesn't contain any. This makes him one of the most grounded and characters among his tier.
- Acts as the bridge between the high-stakes war of the adults and the emotional needs of Harry.
- His dialogue is animated, impulsive, and emotionally charged, revealing a character who acts from deeply felt conviction rather than planning or restraint. This is the reason he belongs in this cluster; not because he is controlled or strategic, but because he acts from internal conviction, not situational pressure.
- His medium SLOAN scores reflect a someone who has been through extreme trauma but has come out the other side with a tempered, realistic perspective
- Personality is driven by emotional conviction and loyalty, giving him agency without discipline and courage without restraint.
- ***NOTE*** difference between him and Harry- Harry reacts because he must while Sirius acts because he believes



Character Descriptions: Severus Snape (SCOAI)

- Personality is defined by extreme internal discipline and emotional suppression
- While sharing the same code as Voldemort, Snape uses these traits for concealment
- Dialogue in the films is minimal, controlled, and precise, reflecting a character who does not externalize emotion but instead channels it into sustained internal commitment
- Unlike reactive characters, Snape does not respond impulsively to events; his actions are guided by long-term intent, even when those actions appear cruel or detached
- What distinguishes Snape within the SCOAI structure is restraint. His agency is quiet rather than demonstrative. He exercises power through consistency and endurance rather than dominance or charisma, maintaining composure under prolonged emotional strain. This makes his influence subtle but persistent
- His personality is built to sustain secrecy, duty, and delayed outcomes



Character Descriptions: Voldemort (SCOAI)

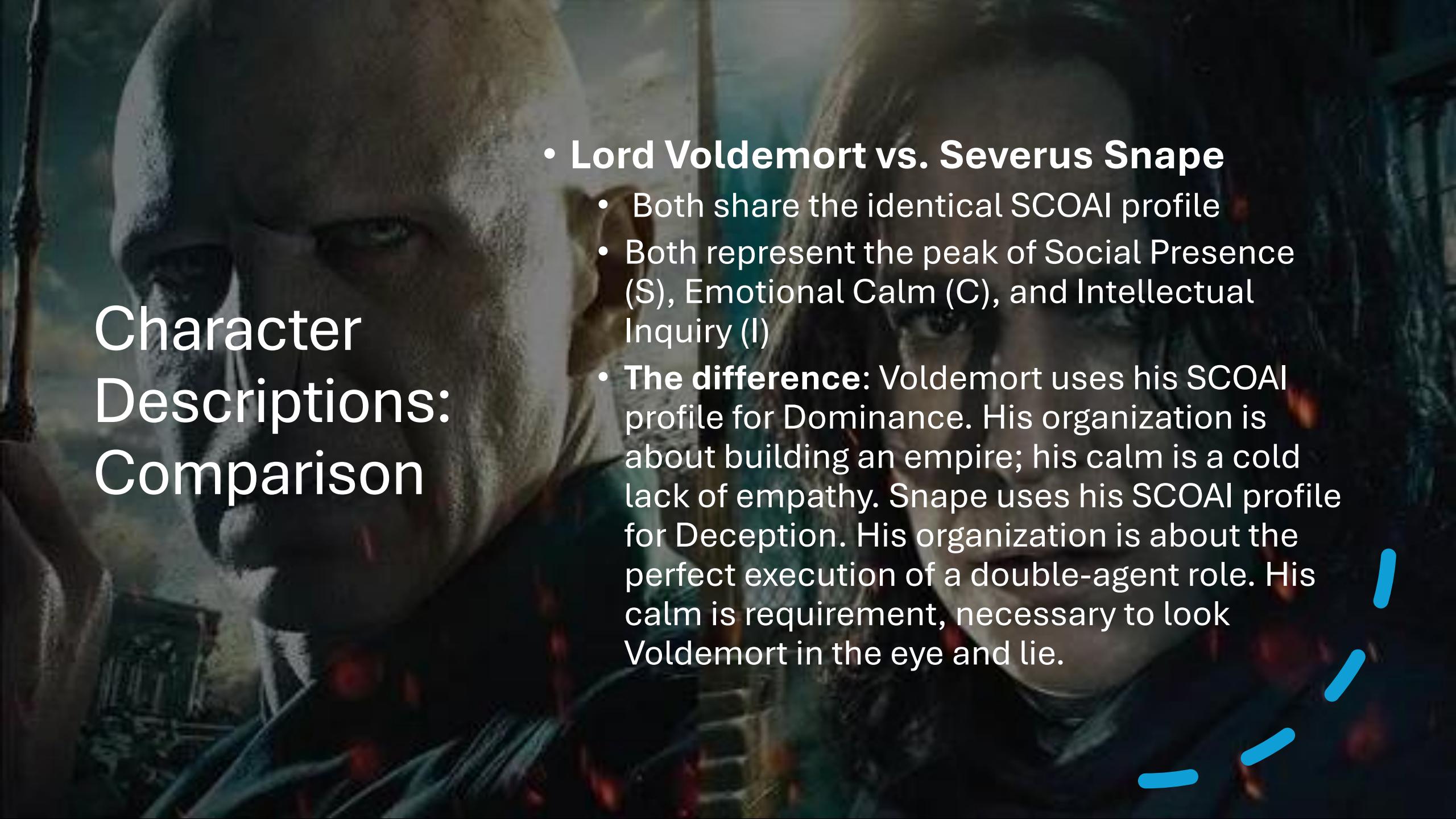
- Voldemort represents the most rigid and self-contained expression of the SCOAI pattern. His dialogue is calm, deliberate, and authoritative, signaling a personality that operates from complete internal certainty
- Extreme Social (S) score reflects his absolute presence; when he speaks, he commands the space.
- Extreme Calm (C) indicates his status as he rarely shouts; he speaks with a terrifying, low-volume detachment.
- His Extreme Accommodating (A) score captures his performative politeness
- He is highly Inquisitive (I) and Organized (O), representing a villain who has researched and planned with academic precision.
- Unlike Bellatrix, he does not display enthusiasm; unlike Snape, he does not suppress internal conflict because conflict is absent.
- In the films, Voldemort acts without hesitation or emotional disruption. His openness manifests as strategic foresight rather than curiosity, and his discipline enables long-term planning without moral friction.
- His emotional control is not restraint but detachment, allowing him to act decisively without relational or ethical interference. He embodies ideological agency stripped of doubt



Character Descriptions: Albus Dumbledore (SCoAI)

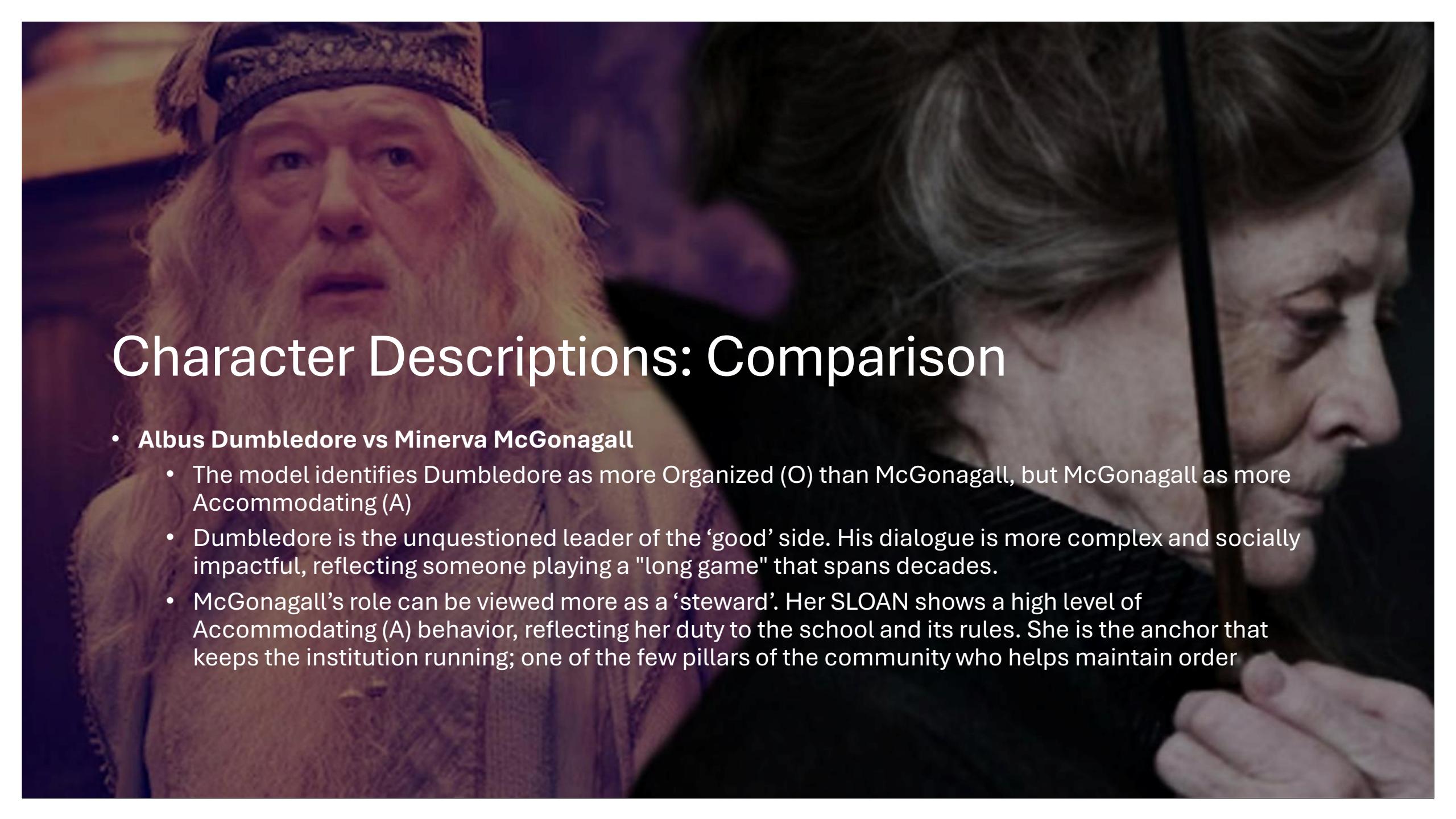
- His dialogue is calm, reflective, and often indirect, suggesting confidence that does not require enforcement
- Nearly a match for Voldemort, with one crucial difference: organized (o). While Voldemort and Snape are rigid and meticulous, Dumbledore's movie dialogue is more flexible.
- His Extreme Inquisitive (I) and Accommodating (A) scores define him as the wise mentor who values curiosity and social harmony over plain dominance.
- Unlike Voldemort, he tolerates ambiguity; unlike Snape, he allows emotional connection; unlike Bellatrix, he does not seek expression through intensity
- Dumbledore's authority emerges from trust and perspective rather than control
- His openness allows him to consider multiple outcomes, while his emotional stability enables patience. He shapes events through guidance rather than imposition, making his ideological power relational rather than absolutist.





Character Descriptions: Comparison

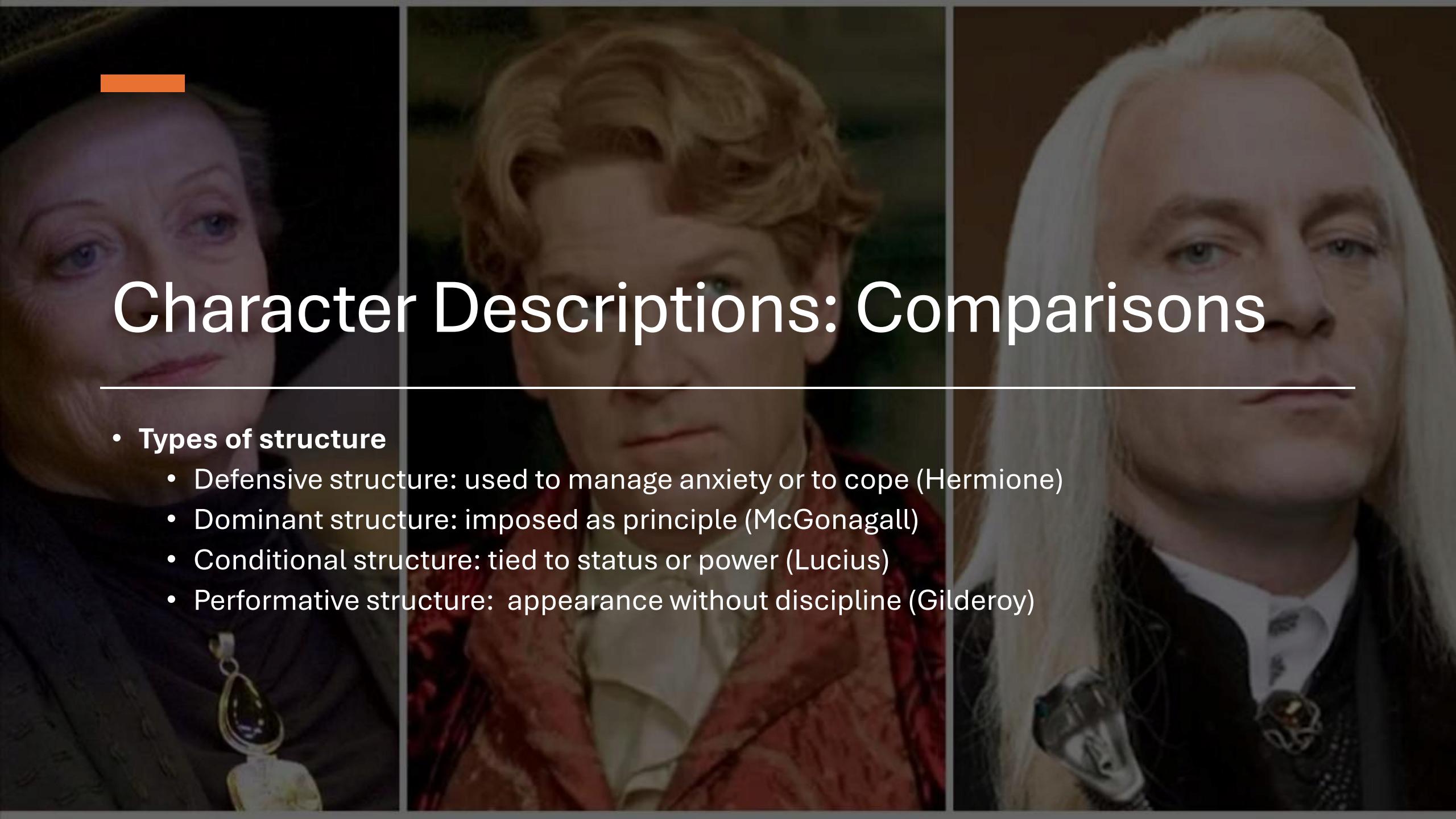
- **Lord Voldemort vs. Severus Snape**
 - Both share the identical SCOAI profile
 - Both represent the peak of Social Presence (S), Emotional Calm (C), and Intellectual Inquiry (I)
 - **The difference:** Voldemort uses his SCOAI profile for Dominance. His organization is about building an empire; his calm is a cold lack of empathy. Snape uses his SCOAI profile for Deception. His organization is about the perfect execution of a double-agent role. His calm is requirement, necessary to look Voldemort in the eye and lie.



Character Descriptions: Comparison

- **Albus Dumbledore vs Minerva McGonagall**

- The model identifies Dumbledore as more Organized (O) than McGonagall, but McGonagall as more Accommodating (A)
- Dumbledore is the unquestioned leader of the ‘good’ side. His dialogue is more complex and socially impactful, reflecting someone playing a "long game" that spans decades.
- McGonagall’s role can be viewed more as a ‘steward’. Her SLOAN shows a high level of Accommodating (A) behavior, reflecting her duty to the school and its rules. She is the anchor that keeps the institution running; one of the few pillars of the community who helps maintain order



Character Descriptions: Comparisons

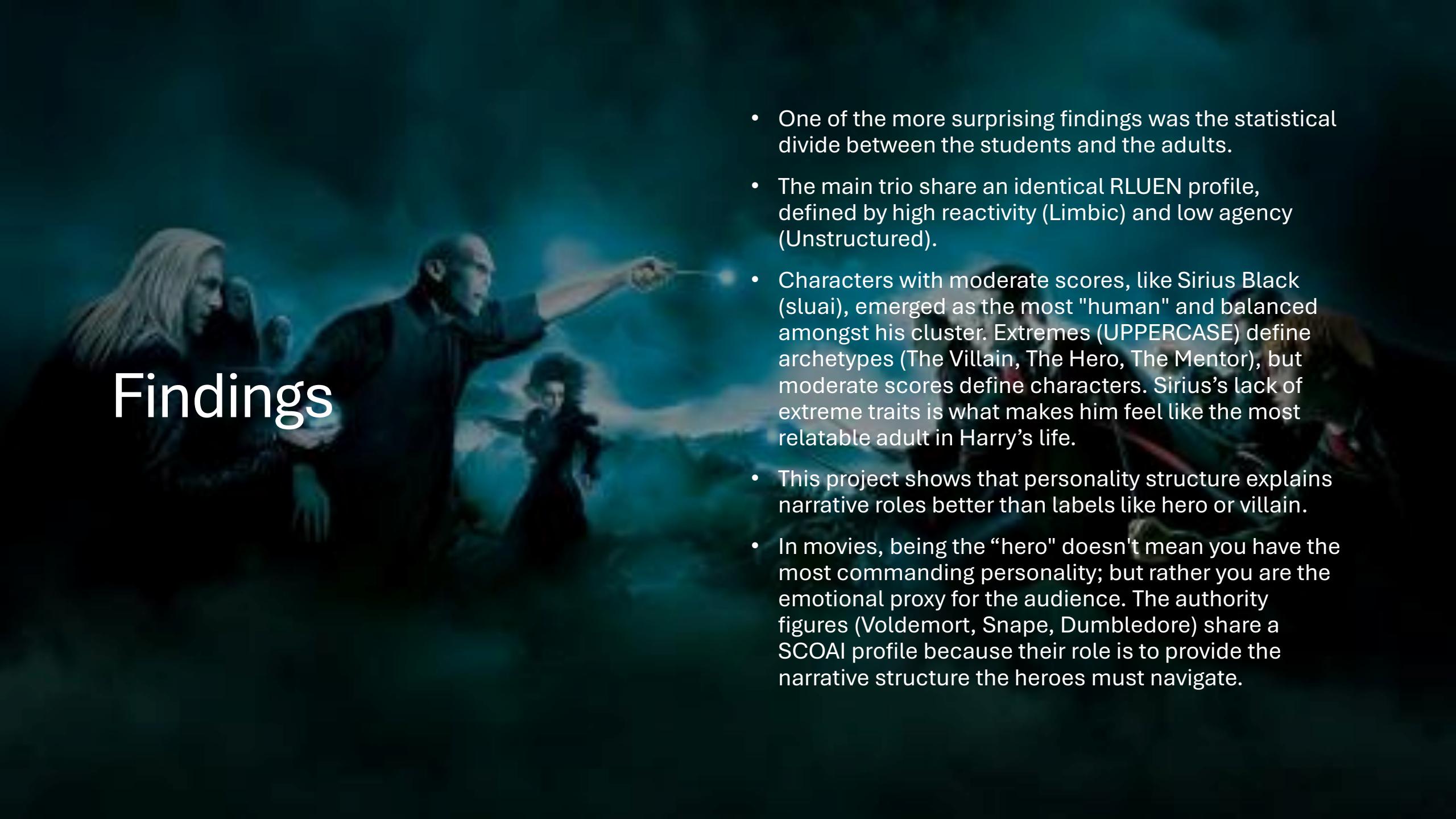
- **Types of structure**

- Defensive structure: used to manage anxiety or to cope (Hermione)
- Dominant structure: imposed as principle (McGonagall)
- Conditional structure: tied to status or power (Lucius)
- Performative structure: appearance without discipline (Gilderoy)



Character Descriptions: Comparisons

- ‘Heroes’ vs the ‘Authority’ : Why are there SLOAN codes so different? (RLUEN VS SCOAI)
- **The Difference:**
 - Reactive vs Agentic behaviors: The main difference between the clusters, found in the dialogue, is how the characters engage with change. Reactive characters in cluster 1 (the main trio), respond emotionally to events and relationships. Agentic characters in cluster 2 (Voldemort, Dumbledore, Snape), initiate the action based on duty or conviction
 - Cluster 2 characters speak in statements; they explain the world. Cluster 1 speaks in questions and exclamations; they experience the world.
 - This might be an intentional trait of characters in cluster 1 from an entertainment perspective, as in fictional stories and adventures, the personality score can be perceived as a reflection of agency. The protagonists score high in Limbic (L) and Unstructured (U) traits because they are our emotional proxies; they feel the wide range of emotions, so we don't have to.

A dark, atmospheric scene from the Harry Potter film. In the foreground, Harry Potter (played by Daniel Radcliffe) looks up at Sirius Black (played by Ewan McGregor), who is casting a spell with his wand. In the background, Remus Lupin (played by Alan Rickman) and other characters are visible in a smoky, moonlit environment.

Findings

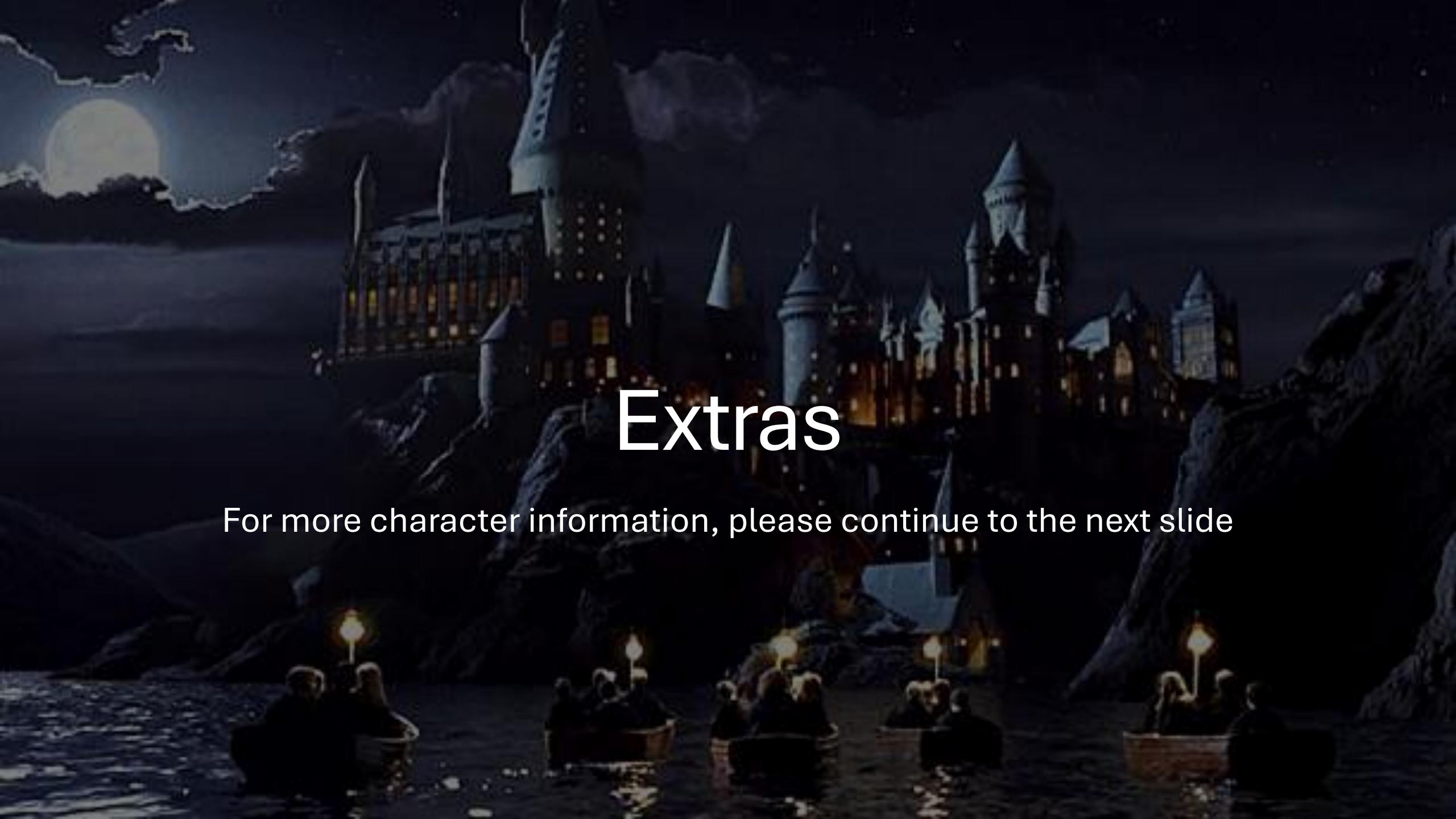
- One of the more surprising findings was the statistical divide between the students and the adults.
- The main trio share an identical RLUEN profile, defined by high reactivity (Limbic) and low agency (Unstructured).
- Characters with moderate scores, like Sirius Black (sluai), emerged as the most "human" and balanced amongst his cluster. Extremes (UPPERCASE) define archetypes (The Villain, The Hero, The Mentor), but moderate scores define characters. Sirius's lack of extreme traits is what makes him feel like the most relatable adult in Harry's life.
- This project shows that personality structure explains narrative roles better than labels like hero or villain.
- In movies, being the "hero" doesn't mean you have the most commanding personality; but rather you are the emotional proxy for the audience. The authority figures (Voldemort, Snape, Dumbledore) share a SCOAI profile because their role is to provide the narrative structure the heroes must navigate.

Conclusion

- Dialogue-based NLP captured patterns of personality expression
- PCA and Clustering helped reveal how characters engage with the world
- SLOAN provides an interpretable bridge between data and narrative
- By modeling personality from dialogue, I noticed that what separates characters isn't morality, but whether they react to change, shape it, or preserve it.
- Similar personalities can produce very different outcomes depending on emotional expression, conviction and context
- I went into this looking for the differences between Gryffindors and Slytherins or 'Good' vs 'Evil'. What was found instead was a map of narrative agency. I discovered that in the Harry Potter films, your personality isn't determined by which House you're in, but by whether you are the one shaping the story or the one surviving it.

Cites

- <https://www.simplypsychology.org/big-five-personality.html>
- <https://personalitynft.com/personality/traits/big-5/sloan/>
- <https://typemyvibe.ai/blog/big-five-guide>



Extras

For more character information, please continue to the next slide

Character Descriptions: Argus Filch (rCOel)

- In the films, he is the institutional shadow, sharing Cluster 3 with Slughorn—both representing the school's old guard from very different angles.
- Defines his identity through enforcement of rules rather than personal authority
- His rCOel code reflects a man who finds order (O) only in the misery of others, obsessively cleaning the same spot on the floor while waiting for a rule to be broken.
- Expresses emotional frustration indirectly through punishment and surveillance
- Represents institutional order stripped of legitimacy and respect
- While he lacks magic, his inquisitive (I) nature is twisted into a surveillance-based obsession with the student body's secrets.



Character Descriptions: Horace Slughorn (sCUEi)

- Motivated by comfort, recognition, and social proximity to power
- A social individual (s) who treats people like ingredients in a potion, carefully selecting the most 'potent' individuals for his Slug Club.
- Dialogue reveals charm paired with avoidance of moral accountability
- His Cluster 3 placement suggests he is more of an opportunist than a leader, preferring the comfort of the sidelines to the heat of the battle.
- Displays a relaxed, unstructured (U) approach to morality, famously 'editing' his own memories to avoid the guilt of his connection to Tom Riddle.
- Prefers influence through networking rather than authority or confrontation
- Organization and ambition surface selectively when personal reputation is at stake



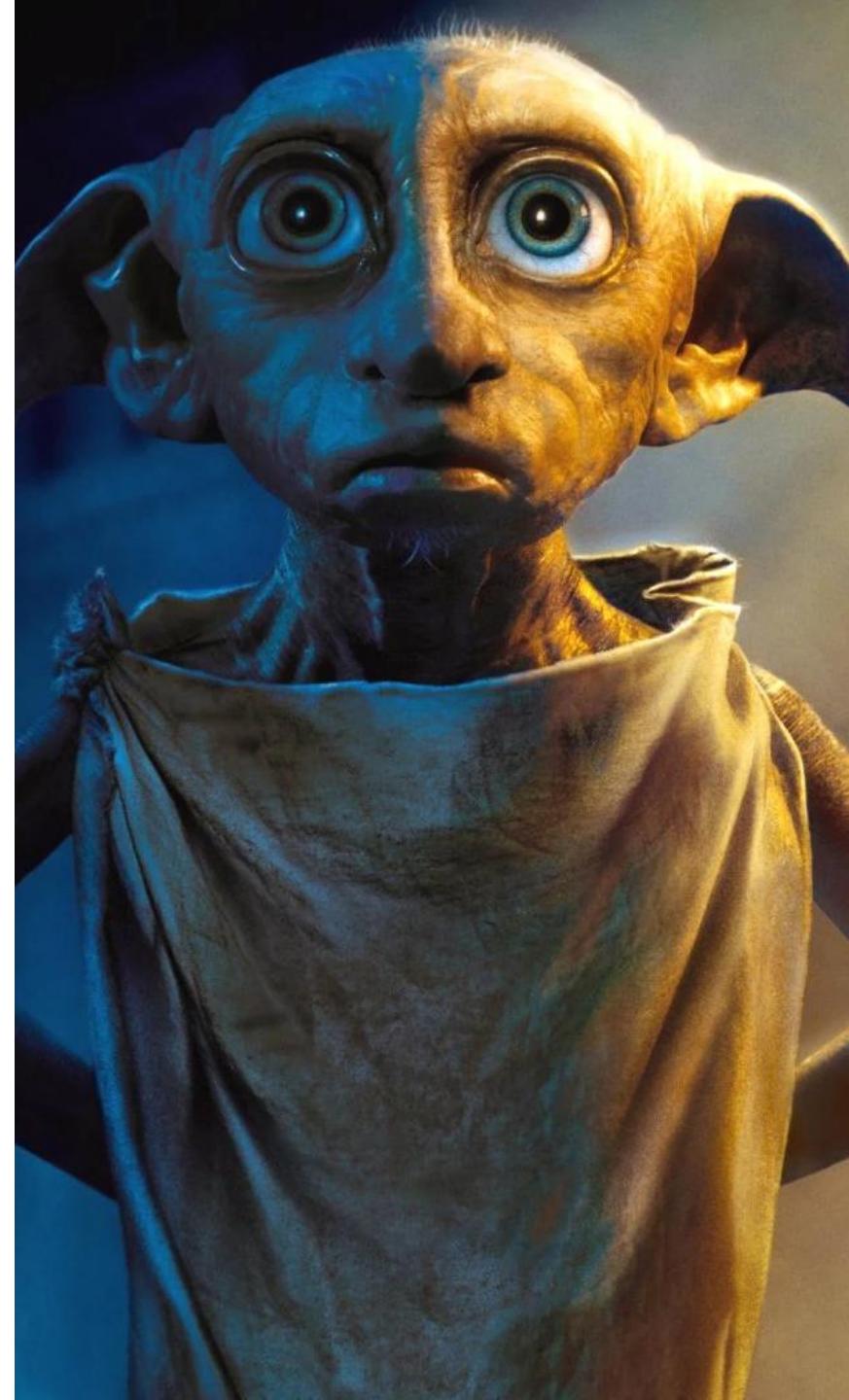
Character Descriptions: Petunia Dursley (SLOAi)

- Fascinatingly separated from Vernon into Cluster 2, Petunia's personality is driven by a limbic (l) jealousy and social (S) anxiety that her husband simply doesn't share.
- Uses rigid household control to suppress perceived disorder
- Emotional expression is tightly constrained and often displaced as hostility
- Uses rigid household control to suppress perceived disorder
- She is the organized (O) architect of the 'perfect' life on Privet Drive, a facade built to hide the deep resentment she feels for her sister Lily.
- In the movies, her accommodating (A) nature is reserved strictly for her son Dudley, while she treats Harry like an unwanted orphan
- Dialogue reflects anxiety toward difference and unresolved resentment



Character Descriptions: Dobby (ScOAi)

- Statistically grouped in Cluster 2 alongside his oppressors, highlighting his high social (S) drive and organized (O) devotion to a cause.
- His movie portrayal emphasizes a selfless, accommodating (A) spirit that drives him to save Harry even when it requires extreme self-injury.
- Dialogue shows intense emotional expression paired with moral clarity
- Willingly endures suffering to protect others
- Unlike other House-elves, his calm (C) resolve under pressure makes him a dangerous wildcard
- Represents the emergence of selfhood through ethical commitment
- Gradually develops independent agency without losing relational devotion



Character Descriptions: Tom Riddle (SCOAI)

- Dialogue is precise, restrained, and strategically manipulative
- He uses his social (S) brilliance to manipulate Dumbledore and Slughorn before ever casting a Dark Curse.
- His SCOAI code is the blueprint of a visionary: perfectly organized (O), emotionally calm (C) during atrocities, and relentlessly inquisitive (I) about the limits of magic.
- Curiosity is driven by desire for control rather than understanding
- Forms relationships instrumentally rather than emotionally
- The films show him as a master of 'The Long Game,' building his Horcruxes with a level of conscientiousness that few wizards have ever matched

Character Descriptions: Dolores Umbridge (SCOAI)

- Displays unwavering confidence in institutional authority
- Dialogue remains calm, patronizing, and emotionally detached from harm caused
- Shares the exact same SCOAI code as characters like Tom Riddle Bellatrix and Snape
- Weaponizes the Ministry's organization (O) to create a polite, social (S) form of torture, masking her egocentric (E) cruelty behind a high-pitched giggle.
- Exhibits no visible self-doubt even under resistance
- Her character is defined by a total lack of empathy
- Represents authoritarianism disguised as civility



Character Descriptions: Sybill Trelawney (SCOAI)

- The **MOST** unexpected member of the SCOAI group
- Dialogue is expressive, speculative, and emotionally charged
- Her high social (S) drama is a defense mechanism against a world that doesn't believe in her 'Inner Eye'.
- Represents openness without disciplined agency
- Emotional volatility alternates with moments of genuine insight
- Maintains an inquisitive (I) obsession with death and tragedy, using the organization (O) of tea leaves and crystal balls to ground her limbic fears.



Character Descriptions: Molly Weasley (sloaN)

- Holds the highest non-curious score, preferring the safety of the home to the dangers of the world.
- Bridges the gap between her adventurous children and the Ministry world, appearing in Cluster 1
- Emotionally expressive with strong relational focus, her dialogue reflects care, worry, and fierce loyalty
- Her limbic (l) worry is her primary motivator, manifesting as the legendary rage she displays when her family is threatened. Is capable of decisive action when in this state.
- Represents authority grounded in connection rather than ideology



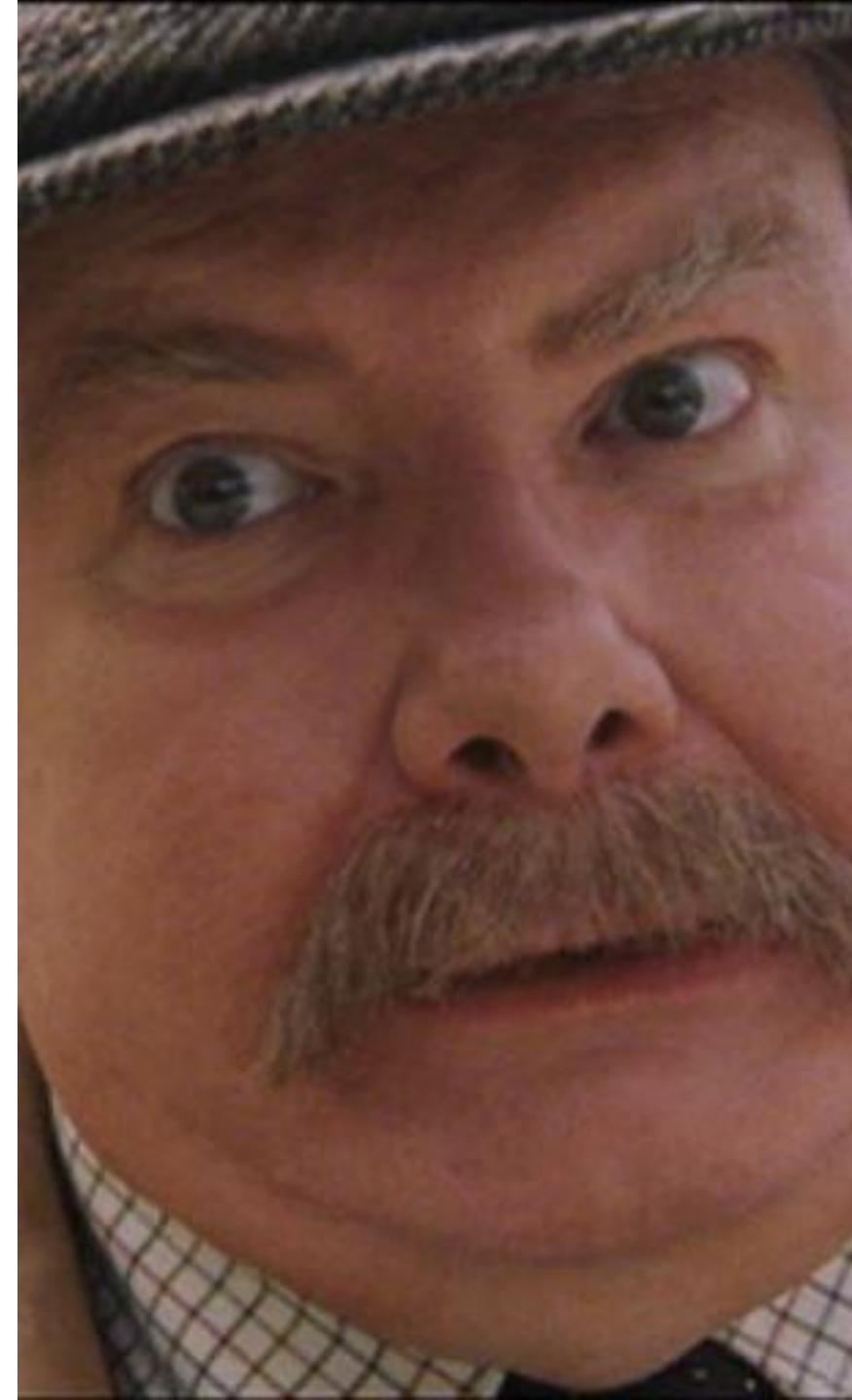
Character Descriptions: Cornelius Fudge (scuen)

- A Cluster 1 outlier who shares the same group as Harry yet uses his social (s) power to actively undermine him to protect his own job.
- Relies heavily on institutional validation and external reassurance
- Dialogue reflects denial when confronted with truth
- His code scuen marks him as the ultimate status-quo seeker: non-curious (n) about Voldemort's return and too egocentric (e) to admit his own mistakes.
- His lack of internal structure (u) leaves him paralyzed when the reality of war finally breaks through his denial.
- Avoids decisive action under threat
- Represents leadership constrained by fear of accountability



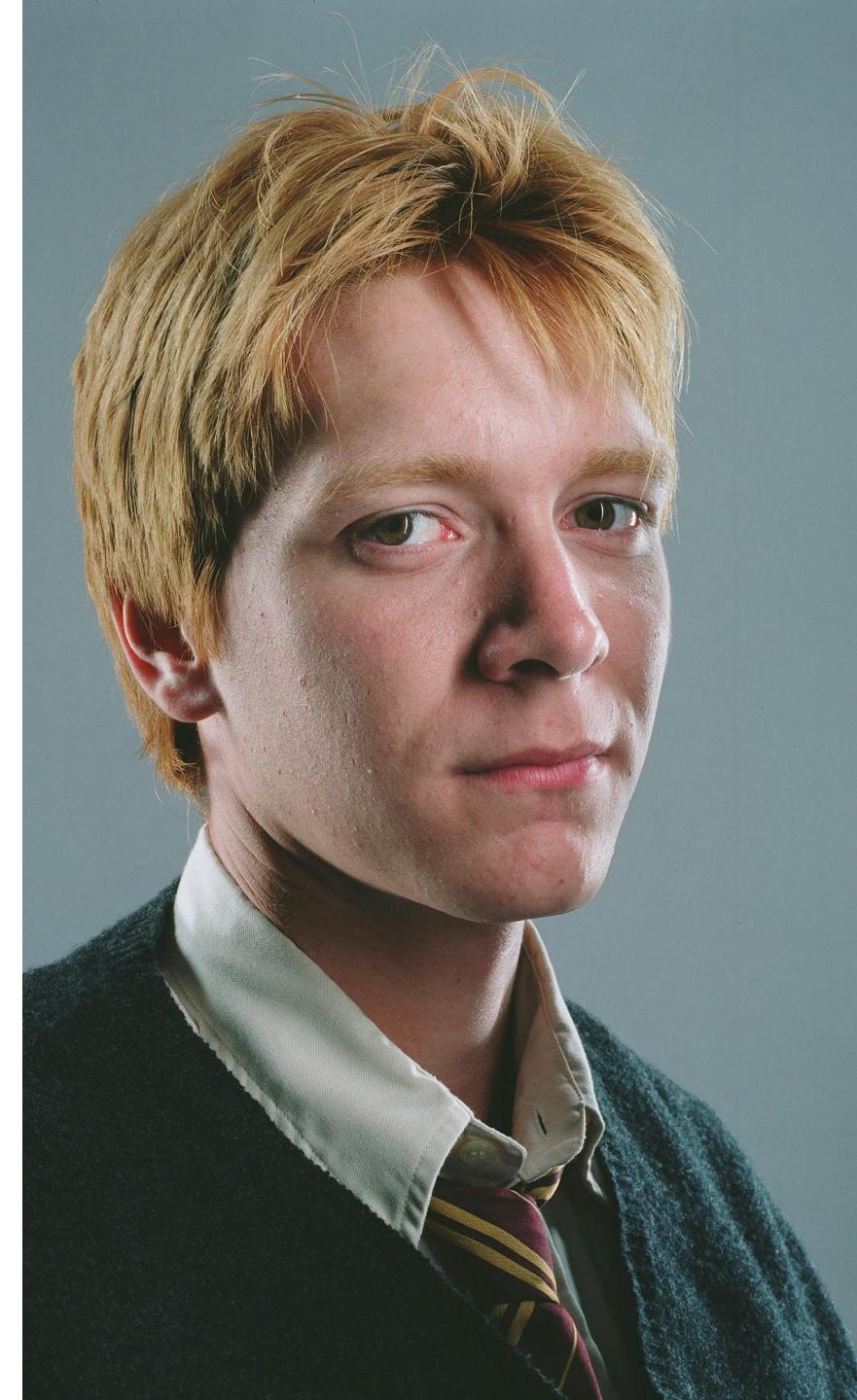
Character Descriptions: Vernon Dursley (rLoai)

- Unlike his wife, Vernon appears in Cluster 1, representing a person whose personality is almost entirely built on 'Normalcy' and traditional (n) values.
- His code 'rLoai' shows a man who is reserved (r) until provoked, at which point his limbic (L) temper takes over in a flurry of purple-faced shouting.
- Emotional expression is externalized through anger
- Uses structure as a tool of suppression rather than order
- Dialogue reflects hostility toward difference and loss of control
- Represents authority rooted in fear rather than legitimacy



Character Descriptions: George Weasley (rLUen)

- While nearly identical to Fred, George's rLUeN code suggests he is a little more reserved (r) and potentially more analytical.
- Emotionally expressive and socially attuned
- Rejects rigid structure in favor of flexibility and creativity
- In the movies, he is usually the one to temper Fred's boldness, though he shares the same unstructured (U) rejection of authority and traditional (N) humor.
- Displays resilience through partnership and play
- Represents emotionally grounded reactivity



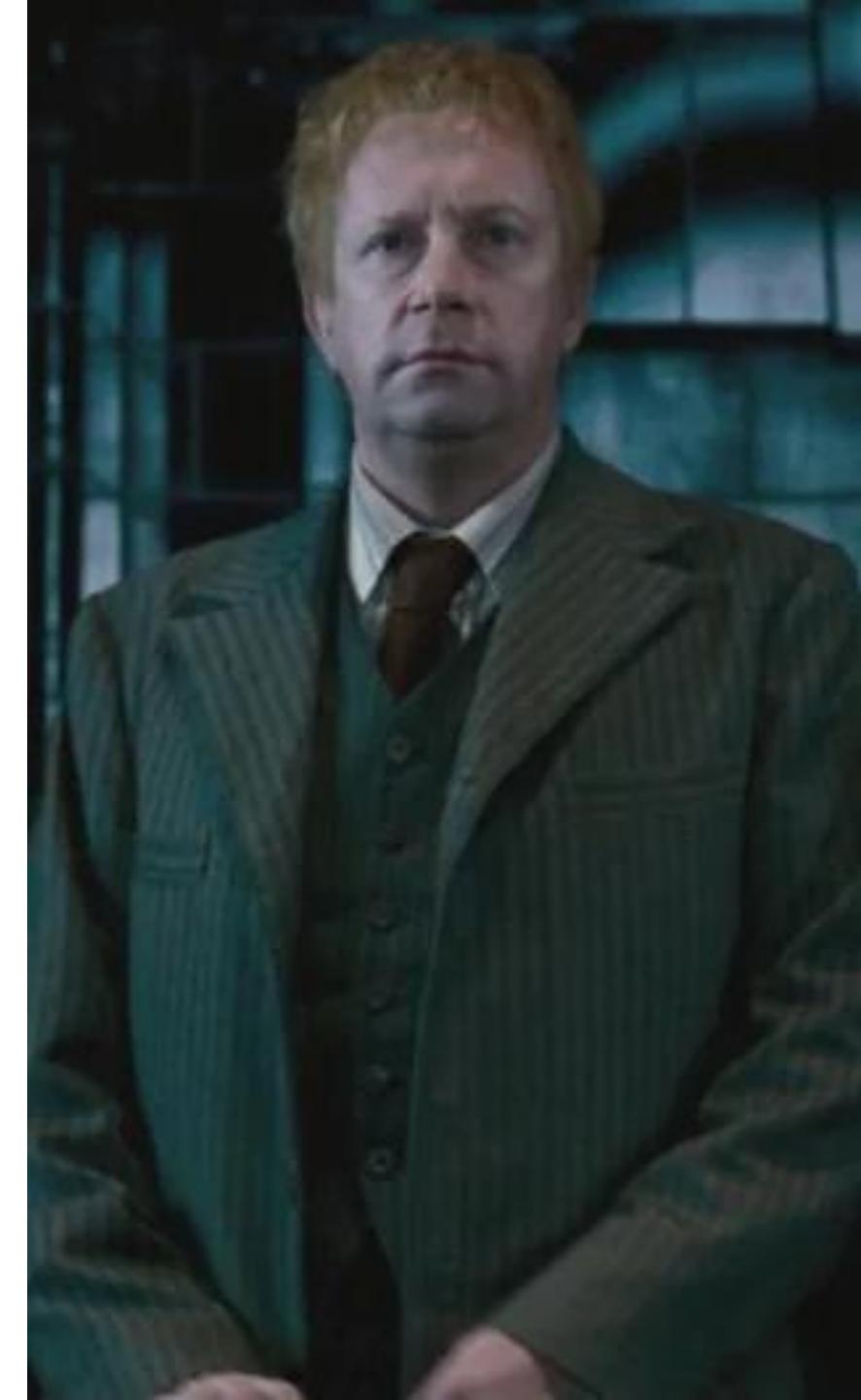
Character Description: Luna Lovegood (RluEn)

- Remains unbothered by the chaos of Cluster 1 because she lives in a world of her own making.
- Dialogue reflects acceptance of unconventional ideas Emotionally steady despite social marginalization
- Exhibits low need for external validation
- Navigates difference without defensiveness
- Her code RluEn shows a reserved (R) soul who is egocentrically (E) confident in her own truth, regardless of other's opinions.
- She provides the emotional balance to Harry, her calm acceptance of death and loss acting as a mirror to his own limbic struggles.
- Represents openness grounded in emotional security



Character Description: Arthur Weasley (RIUEN)

- The quintessential Inquisitive (I) Weasley
- Represents curiosity without threat orientation
- The unstructured (U) father who allows his children to flourish as individuals, standing in stark contrast to the rigid Ministry he works for.
- Structure sometimes seem secondary to exploration
- Emotionally warm and inquisitive
- Dialogue reflects fascination with novelty and understanding
- His Cluster 1 placement reinforces his role as the moral compass of the family, valuing wonder and empathy over power or wealth.



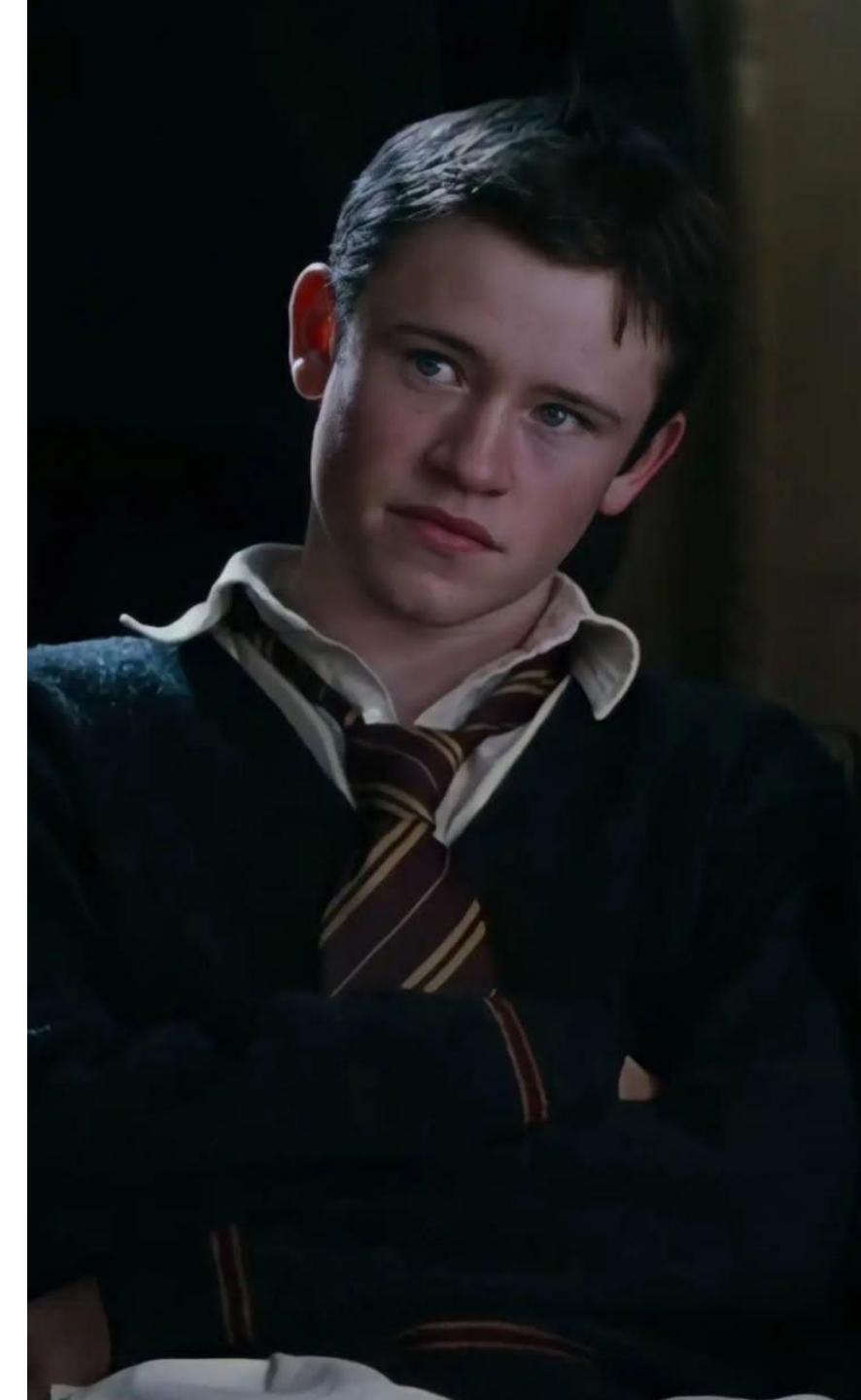
Character Description: Fred Weasley (RLueN)

- The bolder, more limbic (L) twin, Fred's code 'RLueN' reflects a person who lives for the immediate reaction and the loudest laugh.
- Uses humor as both connection and defiance
- Represents joy-driven resistance to constraint
- He is the first to leap into danger and the first to crack a joke, embodying the unstructured (U) spirit of rebellion
- His personality is the 'engine' of the twin's dynamic, driving their social (S) presence even if he technically scores as reserved (R) in private.
- Emotionally expressive with strong social presence



Character Descriptions: Seamus Finnigan (RLuEN)

- A typical 'Limbic' (L) student whose internal tension often manifests physically through his incidental explosions.
- Emotionally expressive and situationally reactive
- His RLuEN code connects him to the traditionalist (N) side of the Gryffindors, often putting him at odds with main trio
- Lacks long-term ideological orientation
- Represents peer-driven reactivity
- Dialogue reflects impulsivity under social pressure
- As a member of Cluster 1, he represents the 'common man' of Hogwarts; social, loyal, but easily swayed by the pressure of the 'Agentic' characters



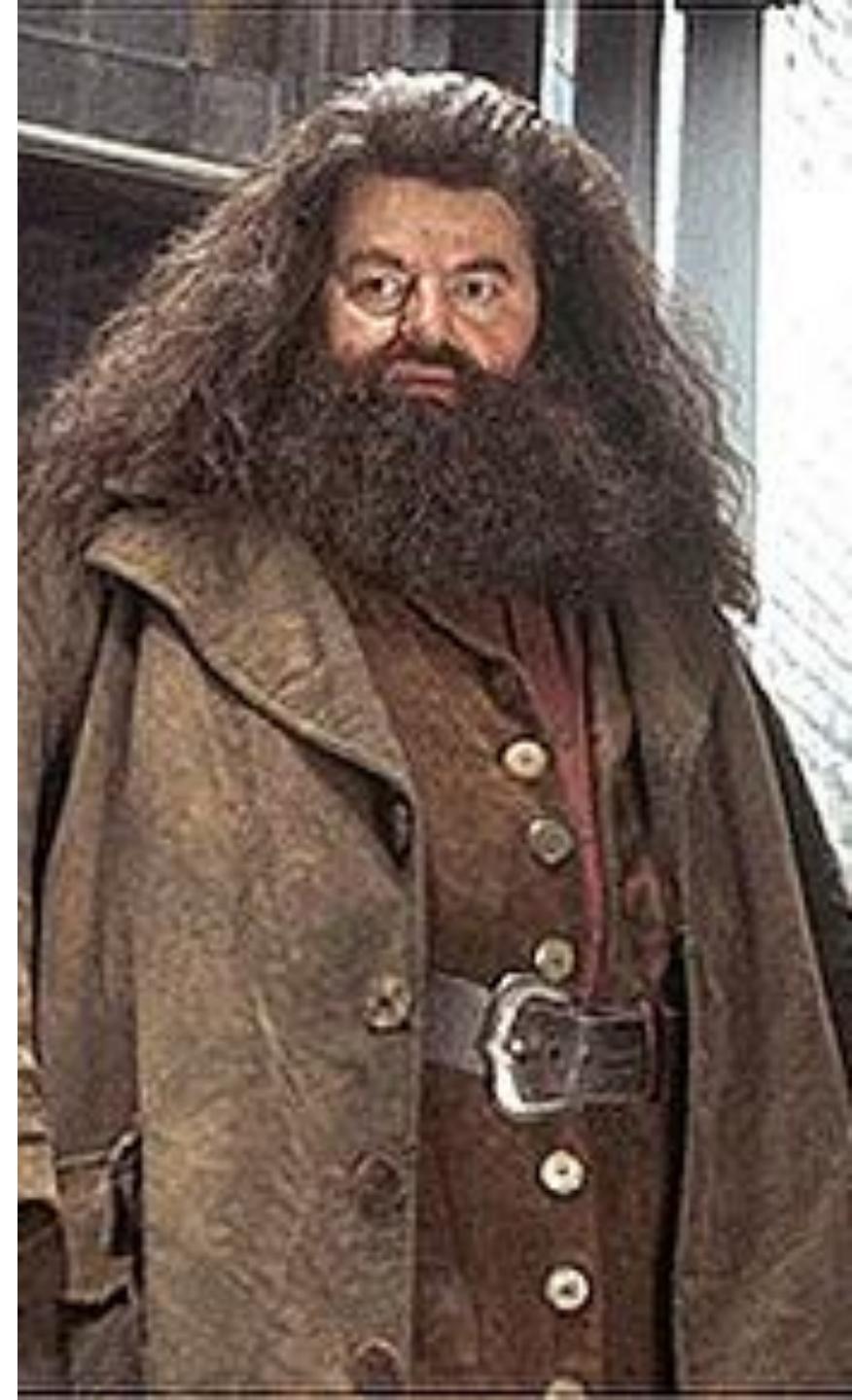
Character Descriptions: Ginny Weasley (RLUeN)

- Starting as a reserved (R) observer, Ginny evolves into a high-limbic (L) leader who is arguably the most fearless of her siblings
- Her 'RLUeN' code is almost identical to George's, showing an unstructured (U) approach to magic that makes her formidable
- Emotionally responsive with growing self-assurance due to her dialogue reflecting an increased assertiveness over time
- Acts decisively without seeking dominance
- Represents maturation within a reactive structure
- She is the bridge of Cluster 1, connecting the playfulness of her brothers with the grit required to fight a war.



Character Descriptions: Rubeus Hagrid (RLUEN)

- A giant whose personality is defined by his limbic (L) heart.
- Emotionally expressive and relationally focused
- Dialogue reflects sincerity and moral simplicity
- He is the most unstructured (U) professor at Hogwarts, teaching through instinct and experience rather than textbooks or exams.
- His loyalty is enacted through action rather than ideology
- Represents emotional steadiness rooted in care
- Him being in Cluster 1 alongside the students, highlights his role as their protector and the heart of the school's 'found family'



Character Descriptions: **Neville Longbottom** (RLUEN)

- Sharing the RLUEN code with Hagrid and the main trio, his journey is one of mastering his limbic (L) fear to become a pillar of courage, completing an arc as grand as Ginny's
- Initially defined by anxiety and self-doubt
- Dialogue reflects gradual internal strengthening
- He is the backbone of Cluster 1, proving that organization (O) isn't necessary for bravery
- Acts decisively when emotionally compelled
- Represents growth within a reactive emotional framework
- One of the more ultimate 'Accommodating' (A) characters, sacrificing his own safety to give others the time they need



The background image shows a large, medieval-style castle at night. The castle has numerous tall, dark spires and towers, some of which are lit from within, casting a warm glow. The sky is dark and filled with clouds, with a bright full moon visible in the upper center. The overall atmosphere is mysterious and dramatic.

THE END

Thank you for your time.