

# Software Project on Deep Learning for Literary Analysis: Diagnosing Mental Disorders in Literary Characters

**Daniil Gurgurov**

Matriculation number: 7033458  
dagu00001@stud.uni-saarland.de

**Nursulu Sagimbayeva**

Matriculation number: 7042653  
nusa00001@stud.uni-saarland.de

**Antonia Wächter**

Matriculation number: 7007542  
anwa00001@stud.uni-saarland.de

**Asmaa Ibrahim**

Matriculation number: 7039914  
asmaibrahiim96@gmail.com

## 1 Introduction

Literature has offered a great opportunity for both authors and readers to dive deeply into the human psyche. Through rich descriptions of the characters' appearance, mental states, and behaviors, authors were able to depict certain mental and behavioral patterns that they could recognize in real life. This was true even before the establishment of psychology as an independent field of study. Over time, the fields of psychology and psychiatry have made progress in understanding these patterns, determining why they can be unhealthy, and categorizing them as various mental disorders.

This project is an attempt to compare the presence of different mental disorders in literary texts across different authors in the same country, and across different countries, within the same time range. In addition to that, the evidence of mental disorders is compared in males against females. This is done by taking the cosine similarity score between character and disorder descriptions.

Our preliminary findings suggest there are notable differences in the distribution of mental disorders in males and females in different countries. However, the focus of the project is rather on demonstrating a possible tool for humanities researchers rather than on making inferences about the connection between mental disorders in literature and reality.

## 2 Project Methodology

In order to achieve the goals of our project, we followed these four main steps:

- data collection,
- data pre-processing and vectorization,
- similarity score calculation,
- comparative analysis.

Furthermore, we developed a webpage to showcase data visualizations, analysis graphs, and an application predicting top-5 most likely disorders given the character description.

## 3 Data Collection

The data collection process was divided into two main parts. Initially, we gathered descriptions of various mental disorders. Following that, we collected descriptions of literary characters.

### 3.1 Mental Disorders Data

In the first part, we acquired the necessary information on mental disorders from the International Classification of Diseases (ICD)<sup>1</sup>, maintained by the World Health Organization - a reliable and credible source. Within the ICD database, we were interested in the 'Mental, behavioral or neurodevelopmental disorders' category. Within this category, we extracted descriptions for 91 disorders by collecting information from the existing sub-categories. This data extraction was facilitated using the ICD API, providing us with names, short descriptions, and essential features of diagnostic requirements for each disorder.

### 3.2 Characters Data

The second part presented more challenges. Initially, our plan was to utilize Named Entity Recognition (NER) to extract character-related data from books by selected authors and then summarize the extracted information. However, despite multiple attempts with various pipelines, the results did not meet our expectations. It became apparent that this problem required its own dedicated project. Hence, we decided to use the available Large Language Models (LLMs), and our final choice fell on ChatGPT. Leveraging this model provided us with well-structured descriptions of equal length

---

<sup>1</sup><https://icd.who.int/en>

for each character, fulfilling our requirements. The primary drawback of using ChatGPT in this context was the challenge of fully relying on the model's output. This issue will be discussed and addressed in a later section. The prompt sentence used to query ChatGPT is as follows: "Give a description of \*name of the character\* from \*author\*'s novel \*novel name\*. Please, mention different aspects of this character: physical, mental, behavioral, social, and other. Also, mention the key actions of this character." This particular prompt was chosen because we tested several options and found that this one consistently gave us good results. An example of a prompt we used and ChatGPT's response are featured in Appendix A.

We started the data collection focusing on the author Dostoyevski since we knew that many of his works feature characters that are likely to have a mental disorder. Completing one dataset initially allowed us to test the performance of our code before collecting the other datasets. After getting good results we continued the data collection and decided on the four countries we wanted to be included in our analysis: Russia, Germany, Great Britain and France. To have more comparable results, we restricted our choice of authors to authors that published novels in the 19th century. The idea behind this was that the language used to describe people with mental disorders may have differed a lot for different time periods.

For each author we created a dataset containing around 40 characters. The entries for each character include the title of the novel, its publishing year, the name of the character, their gender and a description of the character.

This worked quite well for Dostoyevski's characters due to there being enough information on his works in the training data for ChatGPT. However, we soon noticed that for other authors the outputs either included a big amount of inaccurate information or no information at all.

For the German authors the sentence "Use all the information available to you plus the following:" and an already existing character description was added to the prompt. These existing character descriptions came from the following websites:

- <http://literaturlexikon.uni-saarland.de/werke>
- <https://studflix.de/deutsch/thema/literarische-werke-544>

- <https://lektuerehilfe.de/werke>

This approach led to slightly more detailed and reliable character descriptions. Nevertheless, we realize that we are not able to ensure the accuracy of the character descriptions since we have not read all of the original works. In a future continuation of our project these descriptions could be manually corrected to get more reliable results.

Another challenge was to decide on authors that were on the one hand famous enough to have easily accessible information on their characters and on the other hand published enough novels because our goal was to stick with one text type. In the cases of Johann Wolfgang von Goethe and Oscar Wilde we deviated from this goal and incorporated different text types to ensure a sufficient number of characters.

The final selection of authors was Ivan Sergeyevich Turgenev, Fyodor Dostoevsky and Leo Tolstoy as Russian authors and E.T.A Hoffmann, Johann W. Goethe and Theodor Fontane as German authors. We chose Victor Hugo, Alexandre Dumas and George Sand as French authors and Virginia Woolf, Jane Austen and Oscar Wilde as British authors. Oscar Wilde was born in Dublin, Ireland. During his lifetime Ireland officially belonged to the United Kingdom of Great Britain and Ireland and Wilde lived in and published many of his works in London. However, we acknowledge the controversy of classifying him as a British author.

The dataset containing Tolstoy's works encompasses some works published in the early 20th century, while the dataset containing Goethe's works encompasses some works published in the late 18th century. The other datasets only encompass works published in the 19th century. All datasets were saved as csv files.

## 4 Data Pre-processing and Vectorization

To determine how similar characters were to specific mental disorders, we needed to convert our data - character and mental disorder descriptions - into a format suitable for comparison. This involved two main steps: pre-processing and vectorization.

In the pre-processing part, we made sure that the text was consistent and clean. We converted everything to lowercase, removed punctuations, special characters, and stop-words like 'to', 'a', 'the', etc. Then, we performed tokenization by breaking down

the text into individual words and lemmatization by simplifying the words to their base words.

After pre-processing, we used Sentence-BERT (SBERT) to convert the organized texts into meaningful numerical representations known as embeddings. Sentence-BERT, a powerful Natural Language Processing technique, transforms sentences into vectors and, unlike traditional word embeddings, and considers the context of the entire sentence. It employs deep learning models to generate embeddings that are more representative of sentence content, making it ideal for tasks like similarity analysis, clustering, and information retrieval.

The resulting embeddings of our data carried the essence of the characters' descriptions and allowed us to measure their similarity with the mental disorders' descriptions by computing the cosine similarity distance. This similarity calculation served as our method of diagnosing literary characters with mental diseases.

## 5 Similarity Score Calculation

To determine how characters might relate to symptoms of specific disorders, we measured how similar their descriptions were to the symptoms of different disorders. We did this by calculating the similarity between each character's description and the descriptions of various mental disorders. We considered the top three most similar disorders as potential diagnoses for the character.

To calculate the similarity, we used a common Natural Language Processing technique called cosine similarity. Cosine similarity helps us understand how two sets of words resemble each other. It gives a number between -1 and 1, where -1 means completely opposite, 1 means exactly the same, and 0 suggests no similarity. In simpler terms, cosine similarity checks if two descriptions are pointing in a similar direction. This approach helped us assess how characters align with symptoms of various mental disorders in a clear and quantifiable way.

## 6 Comparative Analysis

In this section, we will give a brief comparative analysis of the distribution of mental disorders by country and gender. Note, however, that the data was not checked for validity, and the character samples are not representative of the country, since we took only 3 authors for each country for the purposes of demo.

Overall, the unique count of disorders across all countries was distributed as follows:

- 25 disorders for females;
- 31 disorders for males
- 34 disorders featured in total

Certain disorders (like pyromania or schizotypal disorder) were portrayed uniquely among males, while others (like binge eating disorder) - among females (see Table 1). Fig1 suggests that the distribution of disorders overall follows the power law, with personality disorder being the most common disorder category. Comparing the distribution of disorders between females (Figure 2) and males (Figure 3) suggests that some disorders prevail in male rather than female characters, and vice versa. For example, while Developmental Learning disorder and Conduct-Dissocial disorder ranked as top-3 and top-5 disorders for male characters, respectively, for female characters, the same disorders appeared only once each.

## 7 Visualizations and Application

In order to make our results accessible, we created a database of the datasets collected and deployed our visualizations on GitHub page<sup>2</sup>. Our application is divided into three parts. The main page lists the countries and the writers and directs you to one of the three functions that will be described in detail in this section.

The three different analyses offered by the website are:

- Visualizations of the characters for each author sectioned by the country of the author
- Descriptive statistics for the results obtained
- An application page

### 7.1 Visualizations of characters pages

These are currently four different pages, one for each country. You access the page by choosing a country from the main page. On the page, you can find a visualization for each author. The visualization is a Principle Component Analysis (PCA) for each set of the embeddings of the diseases found in the dataset and the characters in the author's works. As an example, you can find below the graph for

<sup>2</sup><https://d-gurgurov.github.io/projects/project1.html>

Uniquely male disorders	Uniquely female disorders
Factitious disorder imposed on self	Factitious disorder imposed on another
Attention deficit hyperactivity disorder	Binge eating disorder
Secondary obsessive-compulsive or related syndrome	Gender incongruence of childhood
Depersonalization-derealization disorder	
Reactive attachment disorder	
Pyromania	
Agoraphobia	
Bodily distress disorder	
Schizotypal disorder	

Table 1: Unique disorders by gender.

Hoffman’s characters found on the webpage for German authors. Hovering on the points shows the disease or character’s name.

## 7.2 Descriptive statistics page

On this page, the user can choose the country and author of their choice. The visualizations are a comparative analysis of the presence of different disorders in the literature and their distribution among genders (males and females). The code for visualizations is available on our GitHub<sup>3</sup>.

## 7.3 Application page

On this page<sup>4</sup>, the user is asked to give a description of a fictional character, and the output is 5 most likely diagnoses for the description with their similarity scores. Appendix B features an example of the user’s input (taken from a Reddit forum about depression) and the corresponding disorder predictions.

## 8 Limitations and Future Work

In this part we acknowledge and discuss some of our software project’s limitations and propose ideas for future work. It is important to note that even though characters are assigned with mental disorders in our work, our approach is very different to a medical diagnosis and should not be understood as such.

While similarity scores for all characters and mental disorders were calculated, there was no description of a “healthy” person. Future work should investigate whether adding this description leads

to characters that do not exhibit any signs of mental disorders to be classified as healthy. Since for most characters there was not a single disorder that showed significantly higher similarity scores than all other disorders, we used the mental disorders with the five highest scores for each character. A future step could be to identify a threshold for accepting a single mental disorder.

The accuracy of the character descriptions is a crucial requirement for the comparative analysis to be meaningful. In the current state of the project this accuracy can not be assured because of the use of ChatGPT which has a limited knowledge base, a generative nature based on patterns in data and a lack of fact-checking of the outputs. In a future continuation of our project these descriptions could be manually corrected. An idea is to leverage crowdsourcing to gather high-quality descriptions of characters.

We aim to broaden the scope of this work by including a more diverse range of countries, reducing the project’s current focus on Europe. A shortcoming in the selection of authors is the uneven gender distribution: Out of twelve authors, only three are female. Due to legal and economic constraints, fewer women had the opportunity to write and publish novels in the 19th century. In addition to that, the works of many female authors that wrote novels in the 19th century like Amalie Schoppe or Nadezhda Khvoshchinskaya can not be found online. A more detailed dataset should put more work into incorporating female authors. Expanding the datasets with more authors for each country would result in more representative results.

Possible hypotheses that researchers could explore are:

- How represented different disorders are in lit-

<sup>3</sup><https://github.com/nursaltyn/mental-disorders-literary-characters>

<sup>4</sup>[https://d-gurgurov.github.io/projects/mental\\_app.html](https://d-gurgurov.github.io/projects/mental_app.html)

erature? Do some disorders get more attention than others?

- Do the authors tend to ascribe certain disorders (e.g., hysteria or depression) to females, and other disorders (e.g., schizophrenia or sociopathy) to males?
- Do historical events/circumstances influence the depiction of certain disorders in literature (e.g., an increase in the number of characters suffering from post-traumatic stress disorder after the war)?
- Do authors in different countries have a tendency to assign certain disorders to their characters?

In conclusion, while our project has realistic outcomes in characterizing characters with mental disorders, it is important to address some limitations. Enhancing the data quality and promoting inclusivity paves the way to improve the project's utility and reliability in the future.

## A Example of a prompt

**Our prompt:** Give a description of Raskolnikov from Dostoyevsky's novel "Crime and Punishment". Please mention different aspects of this character: physical, mental, behavioral, social, and other. Also, mention the key actions of this character.

**ChatGPT response:** In Fyodor Dostoevsky's novel "Crime and Punishment," Rodion Romanovich Raskolnikov is the complex and tormented protagonist. Here's a description of Raskolnikov based on different aspects:

**Physical:** Raskolnikov is described as a young man in his mid-twenties, of average height and build. He has a pale complexion, dark hair, and sunken eyes. Dostoevsky portrays Raskolnikov's physical appearance as reflecting his inner turmoil and psychological state.

**Mental:** Raskolnikov possesses an exceptional intellect and is deeply introspective. He is characterized by his highly rational and philosophical mind, which often leads him to engage in complex intellectual debates and moral dilemmas. Raskolnikov is prone to periods of intense psychological distress and exhibits signs of mental instability.

**Behavioral:** Raskolnikov's behavior is marked by a mixture of isolation, restlessness, and volatility. He is socially withdrawn and tends to avoid

personal connections. Raskolnikov's actions are driven by a desire for personal greatness and a belief in his own superiority, leading him to commit a heinous crime and subsequently struggle with guilt and paranoia.

**Social:** Raskolnikov's social interactions are strained and often strained. He struggles to connect with others on an emotional level and tends to view them with suspicion or contempt. Raskolnikov's social isolation is reflected in his strained relationships with family, friends, and acquaintances.

**Key Actions:** 1. The central action of Raskolnikov's character is the murder of Alyona Ivanovna, an old pawnbroker. This crime is driven by his theory of the "extraordinary man" who has the right to transgress societal laws for a higher purpose. 2. Raskolnikov's subsequent actions involve his internal struggle with guilt, paranoia, and moral conflict. He experiences mental anguish and inner turmoil as he battles with his conscience and the consequences of his crime. 3. Throughout the novel, Raskolnikov engages in conversations and debates with various characters, exploring themes of morality, suffering, and redemption. These interactions highlight his intellectual prowess and his search for meaning in a morally ambiguous world. 4. Raskolnikov's eventual confession and acceptance of his crime serve as a key action, signifying his moral transformation and pursuit of redemption. He seeks punishment as a way to atone for his actions and ultimately find spiritual and emotional salvation.

Overall, Raskolnikov is portrayed as a complex and conflicted character. His physical appearance, intellectual acuity, volatile behavior, social isolation, and moral struggles contribute to his portrayal as a deeply troubled individual grappling with guilt, redemption, and the consequences of his actions. Raskolnikov's key actions revolve around his crime, his internal battles, and his eventual path towards confession and redemption.

## B Examples of app usage

**User prompt:** When you're looking at something, ever get that feeling that something is off but you can't quite put your finger on it? Now imagine waking up and feeling like everything is wrong, but not really having much of a reason for it. And now, on top of it, imagine that every single little bad thought you've had about yourself is true. You're useless, you'll never amount to anything, you're stupid, all your friends hate you, you're going to

die alone because you don't deserve to be loved. The only person saying those things is your own brain and you can try your hardest to tell yourself it's not true, but you believe all of it because those thoughts are overwhelming and they never stop. Some days are hard. You won't want to get out of bed even though it's a perfectly normal day and there's nothing that you've been dreading. Some days are ok, but you know it's not going to last, so end up spending every single minute wondering when the other shoe will drop. So you don't even get to enjoy it.

**App output:**

Prediction number :1: :

Oppositional defiant disorder (similarity score: 0.5181146036413438 )

Prediction number :2: :

Disinhibited social engagement disorder (similarity score: 0.49928206058013824 )

Prediction number :3: :

Body integrity dysphoria (similarity score: 0.4806559336942201 )

Prediction number :4: :

Depressive disorders (similarity score: 0.4804814149050477 )

Prediction number :5: :

Depersonalization-derealization disorder (similarity score: 0.4755746876498298 )

**C Distribution of disorders by gender**



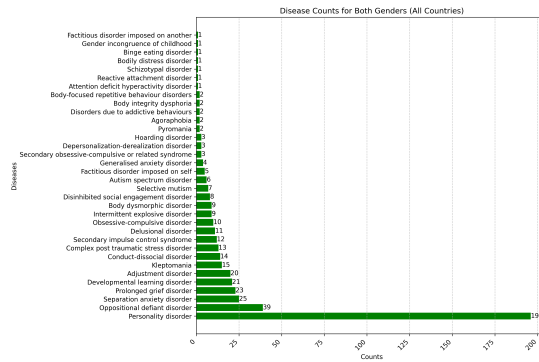


Figure 1: Distribution of disorders (all countries and genders).

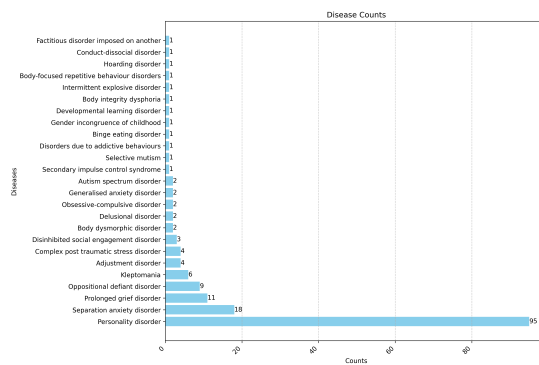


Figure 2: Distribution of disorders (females).

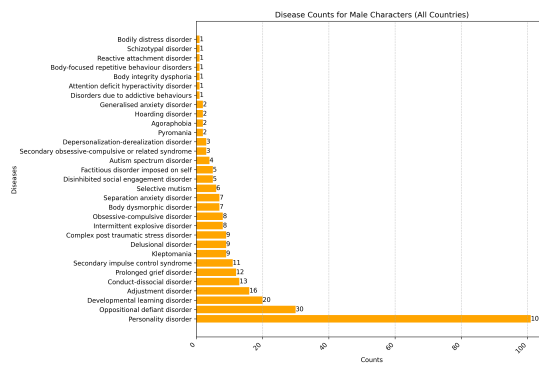


Figure 3: Distribution of disorders (males).