Ps 644: Advanced Topics in AI, Data and Policy

Assignment 2

Abstract

This report employs Latent Dirichlet Allocation (LDA) to uncover latent topics within a corpus of documents centered around the theme of "AI and Creativity." Through the application of LDA, this study unveils hidden patterns, allowing for the identification of key topics and their associated word distributions. The LDA algorithm, applied to a diverse collection of documents ranging from scholarly articles to technical reports, enables a nuanced exploration of the multifaceted discourse on AI and creativity. The report delves into the discovered topics, offering insights into the prevailing themes, connections, and trends within this dynamic field, fostering a deeper understanding of the intricate relationship between AI and the creative domain.

Introduction

In the vast landscape of natural language processing (NLP) and machine learning, understanding the latent structure within large text corpora is a crucial endeavor. Topic modeling, a powerful technique in this domain, seeks to unveil hidden thematic patterns and structures within a collection of documents. Among the various algorithms employed for topic modeling, Latent Dirichlet Allocation (LDA) stands out as a widely used and versatile approach.

Topic Modeling is a statistical modeling technique that identifies abstract topics within a set of documents, allowing researchers and analysts to discern the underlying themes governing the content. This method is particularly valuable in scenarios where the sheer volume of textual data makes manual exploration and categorization impractical. By employing topic modeling, one can automatically uncover the inherent topical structure, providing insights into the major themes and connections within the data.

Latent Dirichlet Allocation (LDA), introduced by David Blei, Andrew Ng, and Michael Jordan in 2003, is a probabilistic generative model that has become a cornerstone in the field of topic modeling. LDA assumes that each document in a corpus is a mix of various topics, and each topic is characterized by a distribution of words. The underlying intuition is that documents exhibit thematic coherence, and topics manifest as sets of words that frequently co-occur within the same context.

The LDA algorithm functions by iteratively assigning words in documents to topics, refining the topic distributions until an optimal representation is achieved. Through this iterative process, LDA uncovers latent topics and their associated word distributions, providing a valuable tool for summarization, content recommendation, and exploratory data analysis.

In this evolving era of information abundance, topic modeling, and LDA, in particular, have found applications across various domains, including academia, industry, and social media analysis. From understanding trends in large datasets to enhancing information retrieval systems, the utility of topic modeling continues to expand, driven by the need to extract meaningful insights from the ever-growing pool of unstructured textual data.

This introduction aims to provide a foundational understanding of topic modeling and the significance of LDA, setting the stage for further exploration into the applications, methodologies, and advancements within this dynamic field of study.

Methodology

This study aims to delve into the nuances of this intersection by employing Latent Dirichlet Allocation (**LDA**) on a comprehensive corpus consisting of **44,724** words spread across multiple documents.

LDA is a powerful probabilistic model commonly used for topic modeling. It assumes that each document in the corpus is a mixture of various topics, and each topic is a mixture of words. By employing LDA, we seek to uncover the latent topics within our corpus and understand the underlying structure that governs the distribution of words across documents.

To implement **LDA**, we leverage the capabilities of the Mallet toolkit, a Java-based package designed for natural language processing and machine learning tasks. Mallet's LDA implementation offers advantages such as efficient processing of large corpora, support for parallelization, and the ability to fine-tune model parameters for optimal topic extraction.

Our corpus, specifically curated around the theme of "AI and Creativity," comprises a diverse set of documents encompassing scholarly articles, blog posts, and technical reports. These documents collectively capture the multifaceted nature of discussions surrounding the synergy between AI and creativity.

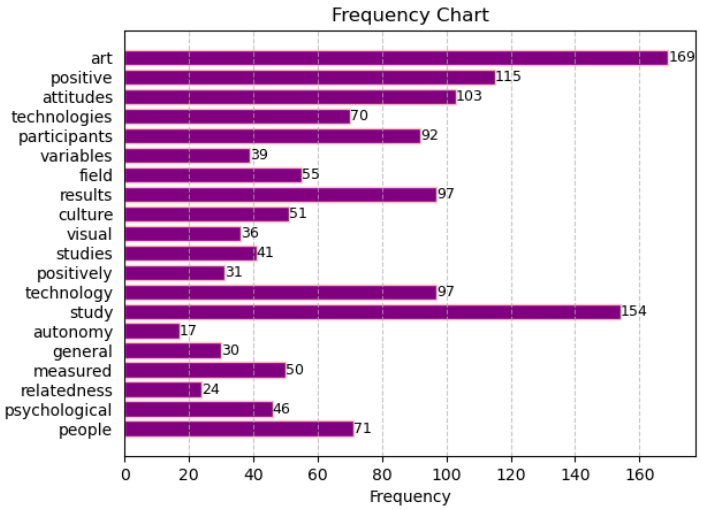
Results

Using MALLET, ten topics were identified, below is the table consisting of the topics and the keywords used to categorize them:

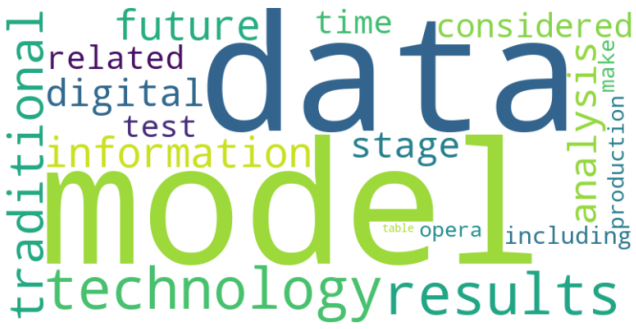
Topic Number	Broad Topic	Keywords
0	Creativity	art positive attitudes technologies participants variables field results culture visual studies positively technology study autonomy general measured relatedness psychological people
1	Digitalization	data technology model analysis table stage future time traditional fig results test considered including related information make production digital opera
2	Empowerment	students creative student genai writing rancie intellectual emancipation outputs authorship process learning tools equality education text teachers prompts images
3	Heritage	cultural heritage intangible year identity research prints development ai-generated perceived ccps products study model print sustainable social impact users aisas
4	Innovation	creativity human creative people thinking fig humans process world methods model education intelligence perspective design potential understanding making co-creation human-ai
5	Intellect	copyright generative law accessed august tools creative works work case guetta originality process author created content original creativity protection creation
6	Intersectionality	haiku human ai-generated human-made hitl beauty poems hotl participants poetry study experience art aigenerated discrimination factors humans hit rate

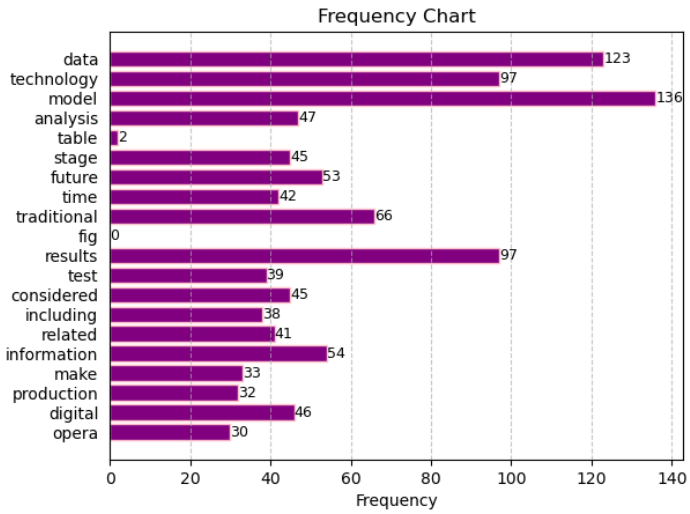
7	Perception	design product authenticity study perceived brand quality attitude consumers fashion effect customization entity process consumer intention clothing ai-designed responses
8	Synthesis	work collaborative human learning collaboration review research article system environment drawing authors humans text september language development literature networks
9	Workforce	employees intelligence behavior innovation organizational creative self-efficacy awareness stara readiness assistants ai-assistant effect study journal ai-enabled management research work artificial

0) Creativity: Positive attitudes towards technology among participants can positively impact the results, fostering a culture of innovation and autonomy. Studies measuring the psychological relatedness of people to technology in the creative field often reveal positive correlations, indicating the significant role of technology in enhancing visual studies and artistic expression

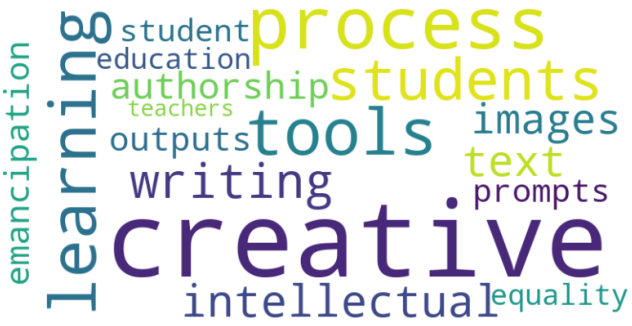


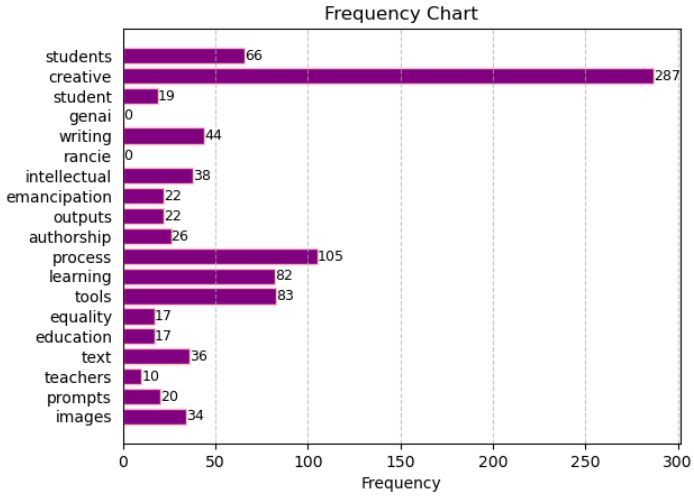
1) Digitalization: The digitalization of creative fields through AI entails harnessing data, technology, and analytical models to propel artistic endeavors into the future. From producing operas to crafting digital artworks, this transformative stage incorporates traditional techniques with cutting-edge tools to generate innovative results and test new creative boundaries.



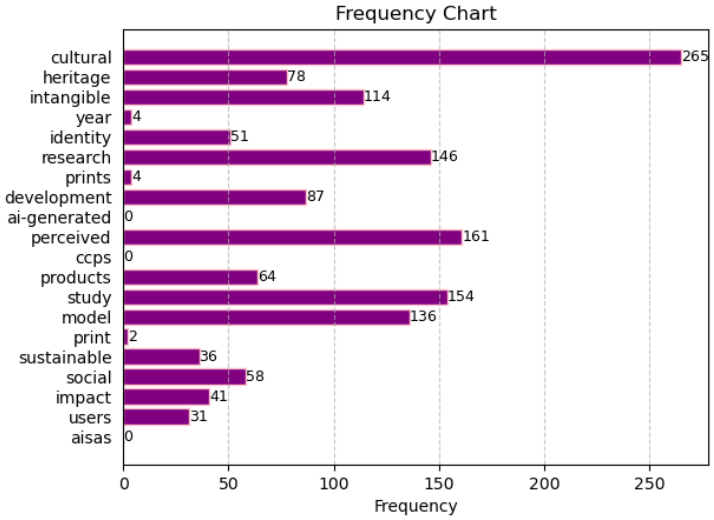


2) Empowerment: AI serves as a catalyst for empowering students by providing innovative learning tools that foster intellectual emancipation. Through AI-generated prompts, images, and text, students can explore their creative potential, redefine authorship, and elevate the process of learning towards greater equality and empowerment.

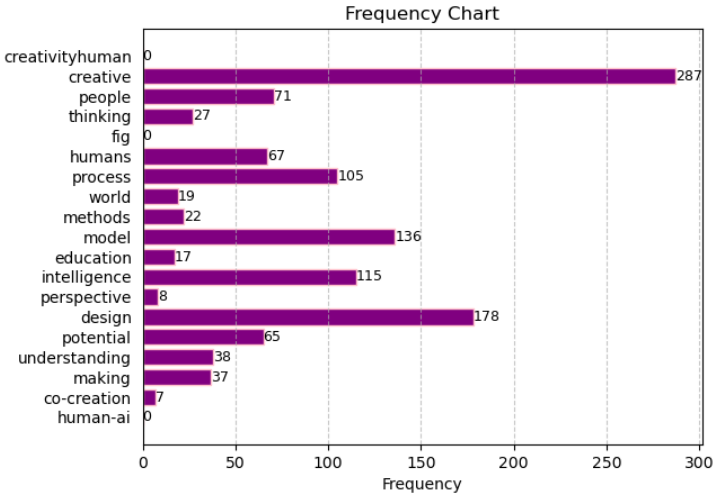




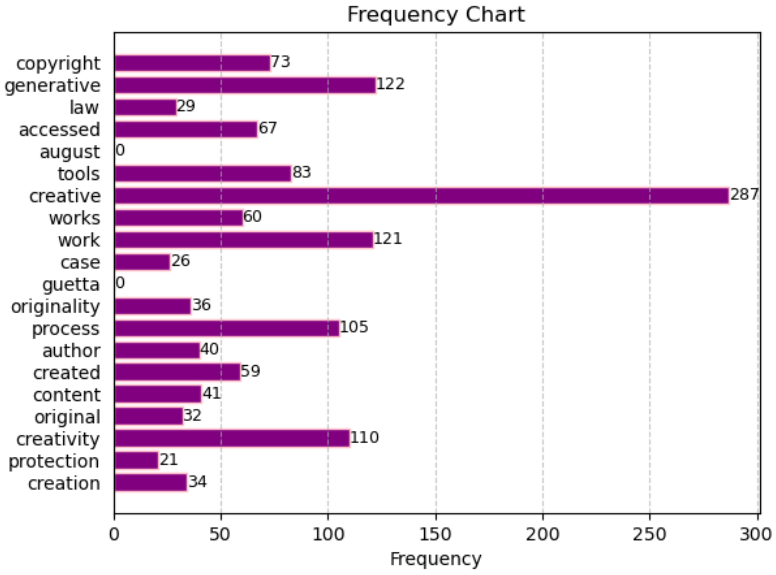
3) Heritage: Cultural heritage takes on a new dimension, blending intangible traditions with cutting-edge technology. Through AI-generated prints and models, heritage preservation intertwines with sustainable development, offering users a bridge to the past while shaping perceptions and identities for years to come.



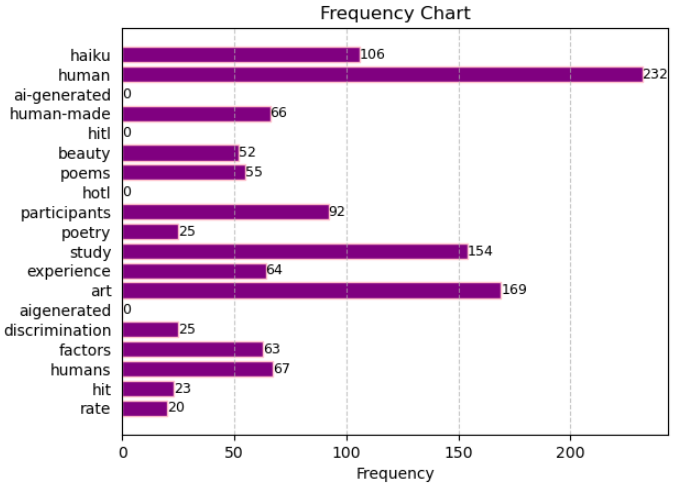
4) Innovation: Collaboration between human creative thinkers and AI models opens new perspectives and methods for design and problem-solving. This co-creation process harnesses the potential of AI to augment human understanding and enrich the creative process, offering a new paradigm in education and creative thinking.



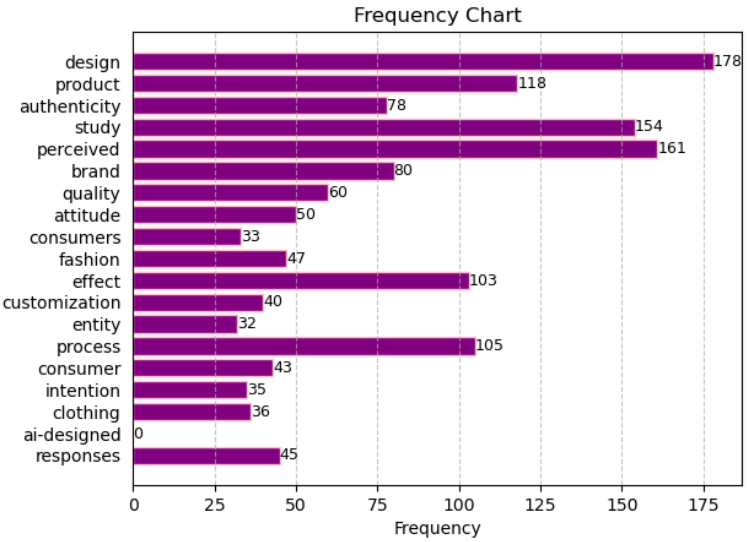
5) Intellect: The concept of intellect intertwines with questions of originality and copyright protection. As AI-generated tools advance, challenges regarding the authorship and originality of creative works, such as the case with David Guetta's AI-assisted music creation, highlight the evolving landscape of intellectual property law and the need for updated frameworks to safeguard creative processes and content.



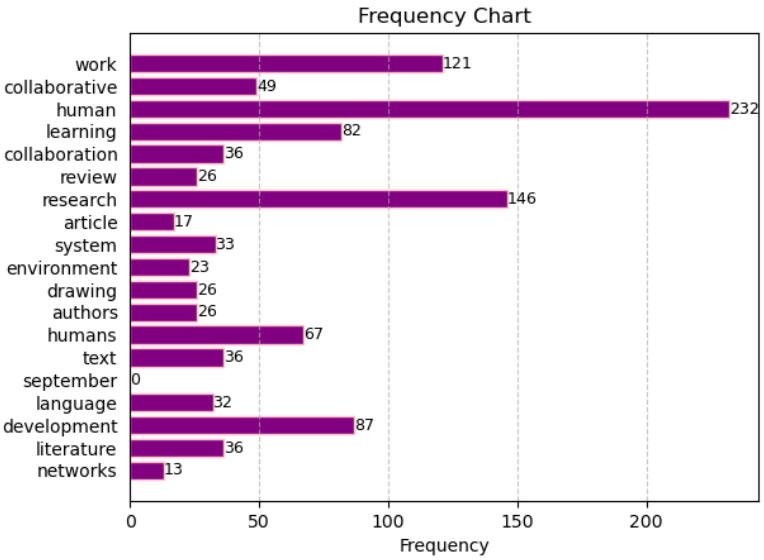
6) Intersectionality: Haiku crafted by both human and AI voices meld, exploring beauty and human experiences. Poetry studies delve into the hit rate of AI-generated and human-made art, illuminating discrimination factors faced by participants in the creative process.



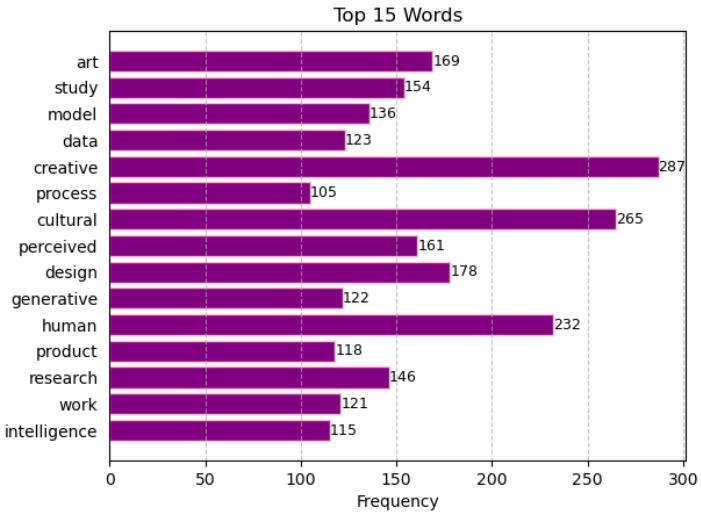
7) Perception: Perception plays a pivotal role in shaping consumer attitudes towards AI-designed products. Studies suggest that the perceived authenticity and quality of a brand's offerings can significantly influence consumer intentions and attitudes, particularly in fashion where customization and design processes are increasingly influenced by AI-generated responses.



8) Synthesis: Synthesis emerges as a pivotal process, where algorithms work collaboratively with human creators to explore new artistic frontiers. Through collaboration, these systems analyze vast swathes of literature, review articles, and research texts, synthesizing insights to augment human creativity and accelerate innovation in diverse creative environments.



9) Workforce: The workforce undergoes a transformation as employees harness AI-enabled assistants to augment their intelligence and enhance creative innovation. Studies in journals explore the effects of AI on organizational behavior, highlighting the importance of readiness and self-efficacy among employees to adapt to AI-enabled management and artificial intelligence-driven workflows.



Top 15 most occurring words in teh corpus were: "art", "study", "model", "data", "creative", "process", "cultural", "perceived", "design", "generative", "human", "product", "research", "work", "intelligence"