

INTRODUCTION

With the rapid advancement of AI and ML technology, **deepfakes** have blurred the line between reality and fabrication. While deepfakes have often been employed for humorous and entertaining purposes, they have always harbored the potential for disruption. Unfortunately some bad actors have exploited this potential and wreaked havoc.

Our study delves into analyzing the deepfake landscape by examining posts and conversations on **Reddit**. We focus on who has been targeted, understanding the tools used to create deepfakes, and capturing the sentiments expressed by the community.

MOTIVATION

- Understanding the impact deepfakes have on society, evaluating its potential for **future disruption**.
- Explore and discover the wide **variety of tools** utilized in the creation of deepfakes, and to assess the potential threats posed by each tool.
- Deepfakes have been polarizing since its inception. We aim to investigate how **public opinion** regarding deepfakes has evolved over the years.

APPROACH

- Crawled **SFW** deepfake-related posts from Reddit using **PRAW API**. Performed text preprocessing with NLTK and spaCy to prepare data for analysis.
- Leveraged Google's Gemini **LLM** via PaLM API to categorize posts and comments into predefined categories which aided further analysis.
- Used spaCy's NER to identify individuals, then created a Python script to count mentions of their name in relevant posts, producing **Fig 1**.
- Applied VADER model to assess sentiment in comments, producing **Fig 2**.
- Conducted **Topic Modeling** with BERTopic, generating interpretable topics. Grouped similar topics into larger themes, shown in **Fig 5**.
- Analyzed posts about deepfake creation tools, identifying tools emerging each year and their frequency. This data was used to create **Figs 3 and 4**.

BENEFACTORS

- **Social Media Platforms:** Develop robust content moderation strategies to curb the spread of harmful deepfake content.
- **Law Enforcement Agencies:** Utilize insights to combat illicit use of deepfake technology for fraudulent activities or misinformation campaigns.
- **Researchers:** Enhance detection capabilities by studying tools and methodologies employed in creating deepfakes.

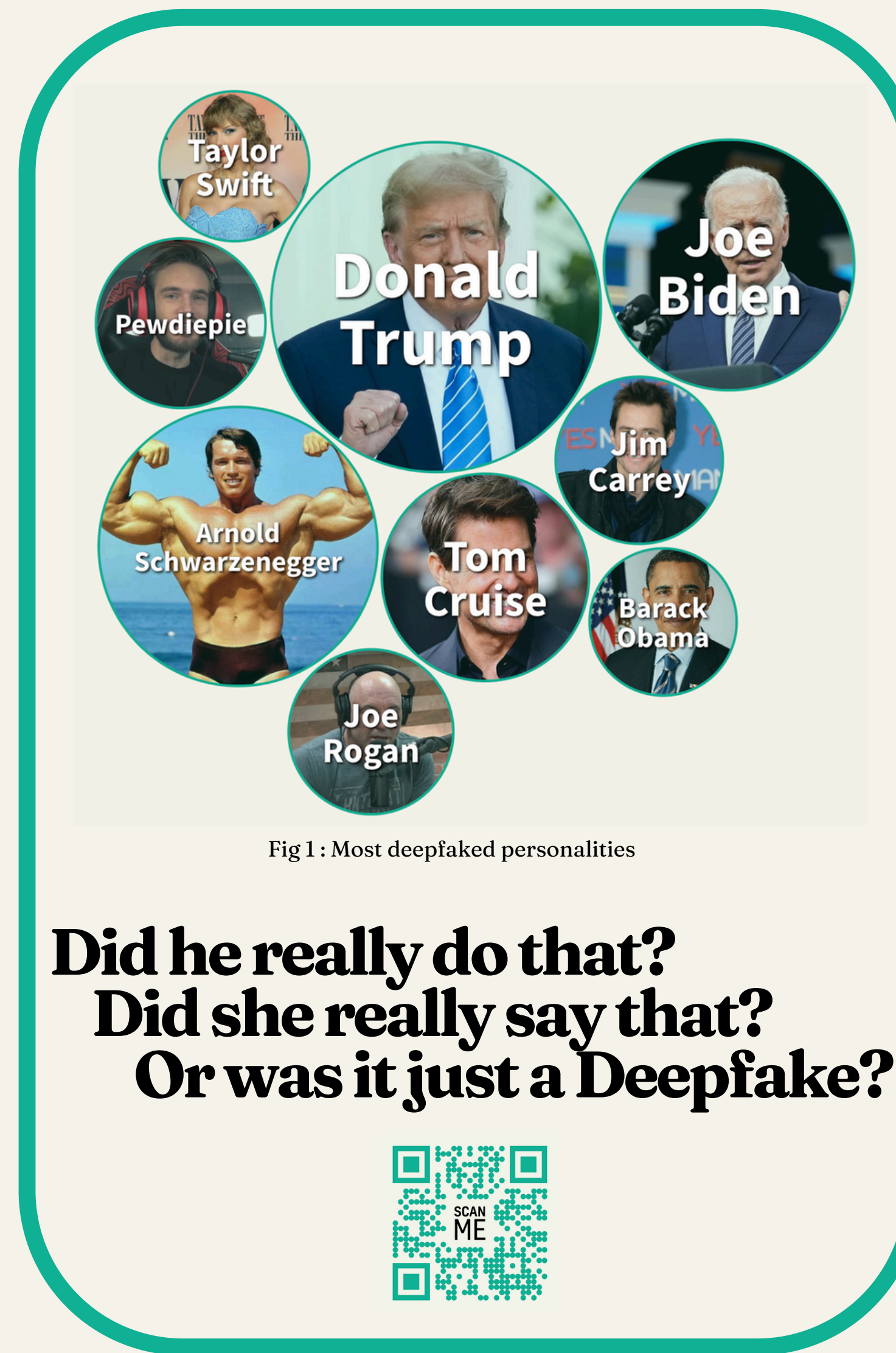


Fig 1 : Most deepfaked personalities

RESEARCH QUESTIONS

1. What are the prominent latent topics and themes present within the deepfake-related discussions on Reddit during this study period?
2. What individuals, groups, or entities have been the primary targets of deepfakes and what were the tools used to create them?
3. How does public sentiment towards deepfake content vary from Jan 2018 to Feb 2024?

RESULTS

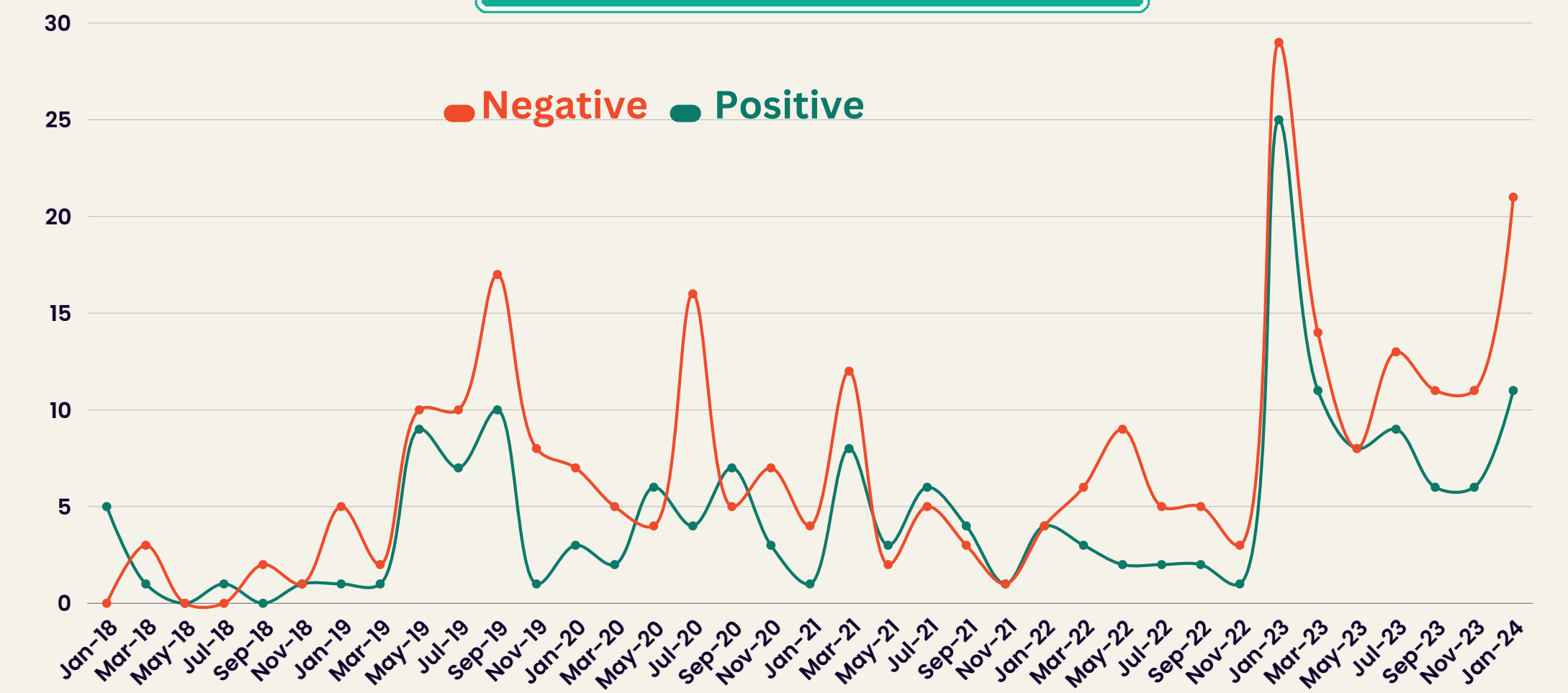


Fig 2 : How has the community perceived deepfake?

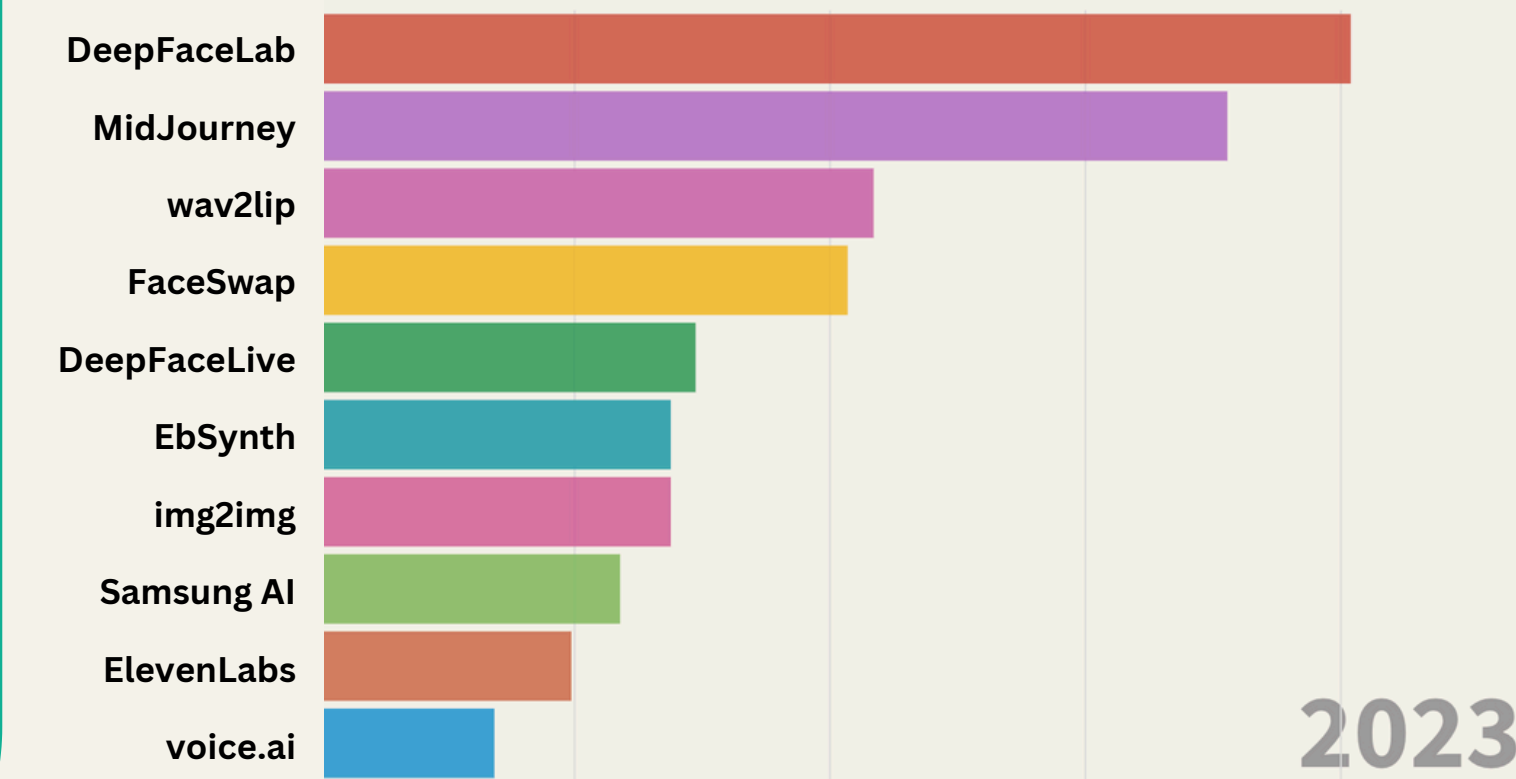


Fig 3 : Popular Deepfake Tools

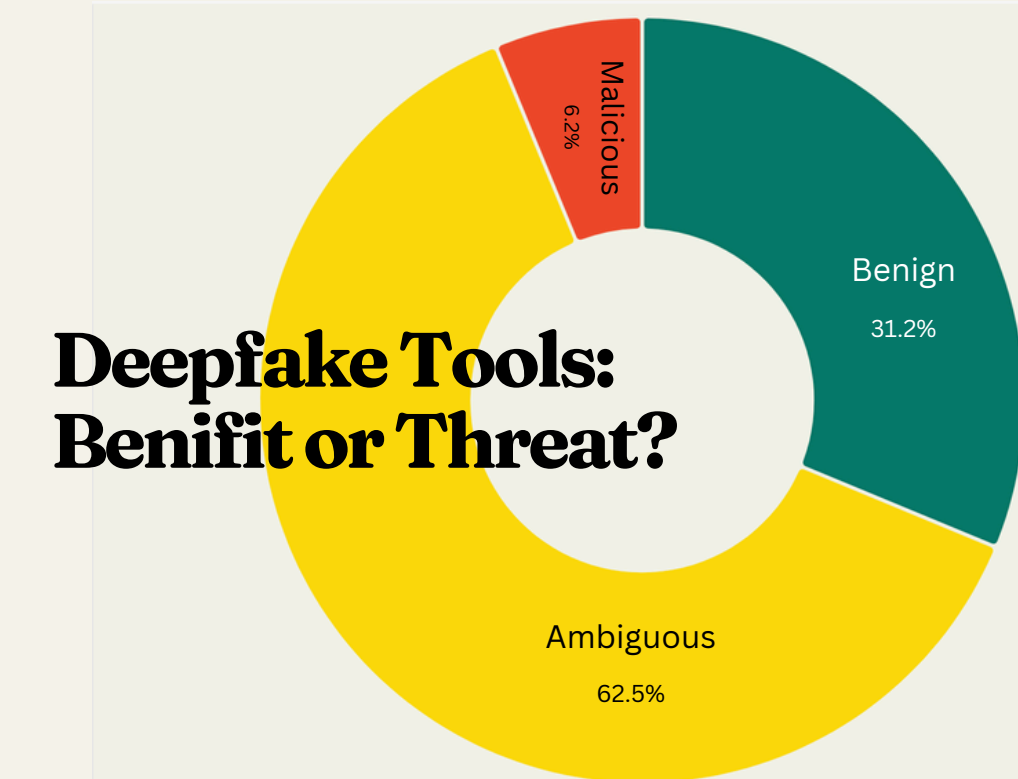


Fig 4 : Deepfake Tools Intent analysis

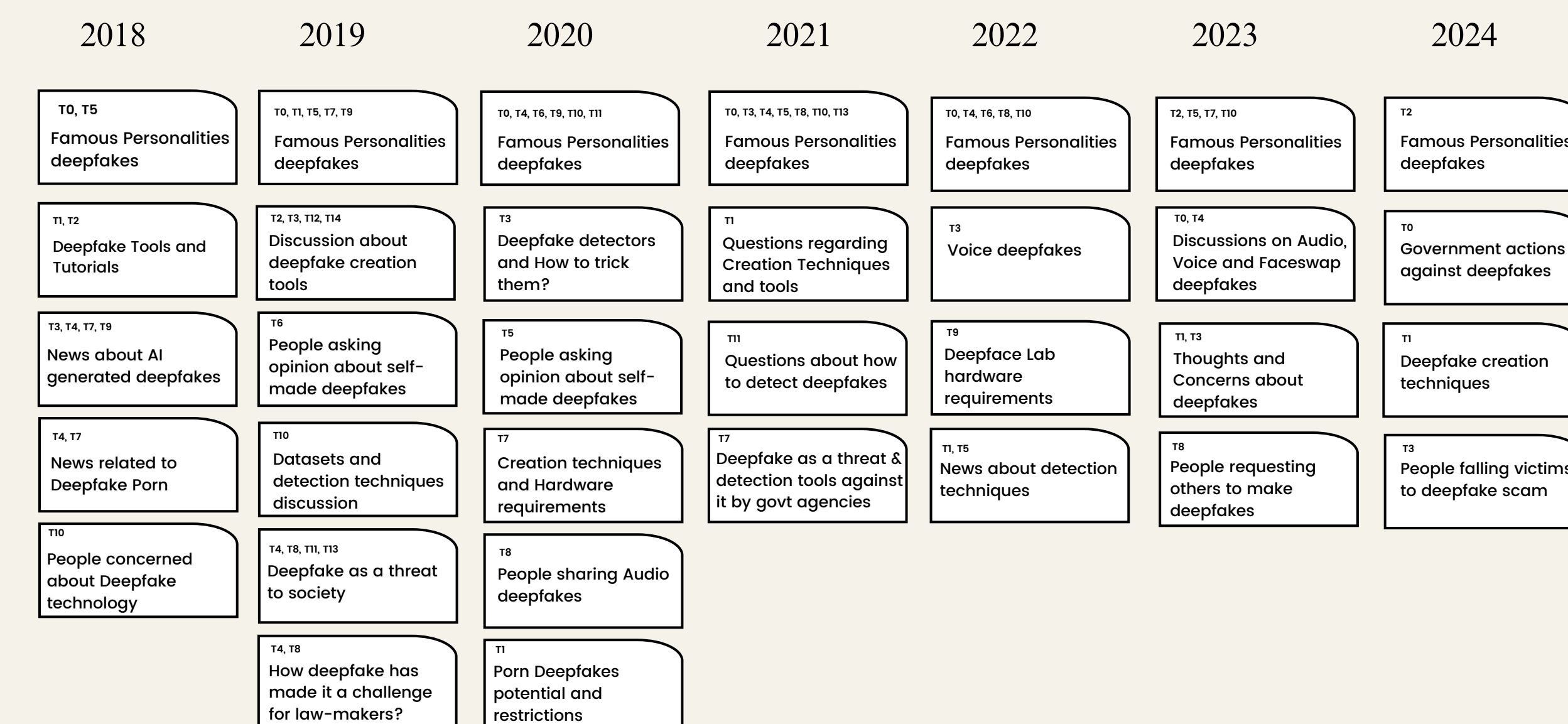


Fig 5 : Topic Modeling results for the study period