1. Biomedical text mining: In the BiomedicalTextMining.ipynb tutorial, you will use PubMed e-utilities to search and download publications about COVID-19 from PubMed. Then, you will use PubTator to retrieve several types of named entities from these abstracts (e.g., diseases, chemicals, genes) and visualize them in several ways to better understand the content of these publications.
2. Social media text mining: In Kyrie’s recent [CSCW paper](https://arxiv.org/pdf/2311.10990.pdf)[[1]](#footnote-1), they unpacked ethical concerns around non-fungible tokens (NFTs), which are decentralized digital tokens to represent the unique ownership of items. Social media analysis was a key part of this mixed-methods study. The scraping and analysis processes are documented on [GitHub](https://github.com/kyriezoe/NFT-cscw-2023). They collected tweets with keywords NFT + {trust, transparent, transparency, racism, morality, moral, inclusivity, inclusive, fairness, ethics, ethical, equity, equality, diversity, diverse, discrimination, bias}. You can replicate the analysis process by running the following jupyter notebook: <https://github.com/PlevanTem/NFT-cscw-2023/blob/main/code/analyze.ipynb>. Key steps include (1) data cleaning, (2) identifying key opinion leaders (KOLs), (3) calculating word frequency, (4) co-occurrence and network analysis, (5) topic modeling, and (6) a comparison between KoLs and public users.

1. Xiao, Y., Deng, B., Chen, S., Zhou, K. Z., LC, R., Zhang, L., & Tong, X. (2024). "Centralized or Decentralized?": Concerns and Value Judgments of Stakeholders in the Non-Fungible Tokens (NFTs) Market. In *27th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW)*. [↑](#footnote-ref-1)