





King Abdullah University for Science and Technology
KAUST Academy
Duke University
DATA+ program

Ethical Consumption Before Capitalism

Raghad Aljohani
Project Leads: Dr. Astrid Giugni, Dr. Jessica Hines, Dr. Nitin Nuthra.

Project Objectives and Goals

This project investigates the ethical dimensions of luxury consumption in Early Modern England by examining how goods—enabled by colonial expansion, global trade, and exploitative labor systems—were represented and morally justified in period discourse. It bridges historical inquiry with computational tools to better understand how ethical debates around trade, consumerism, race, gender, and religion evolved between the late Middle Ages and Renaissance.
Key goals include:

- Analyzing Early Modern English texts for attitudes toward global trade and consumption.
- Training and applying a fine-tuned BART language model to detect ethical themes in historical narratives.
- Creating interactive visualizations and spatial mappings to track the evolution of these discourses.

Ethical Considerations in Trade and Consumption
In Early Modern England, the rise of luxury goods like sugar, tobacco, and exotic dyes, often produced through colonial exploitation and forced labor, sparked moral debate. Economic opportunity clashed with ethical concerns, as merchants pursued profit under the guise of spiritual gain. Propaganda masked the violence behind trade, raising questions still relevant today about the morality of wealth built on exploitation.

Student Learning Outcomes

Participants engage with:

- Fine-tuning Large Language Models (LLMs), especially BART, on historical texts.
- Investigating how ethical language in consumer discourse intersects with gender, race, and religion.
- Developing interactive tools like RShiny apps and geospatial mappings to visualize temporal and spatial changes in ethical discourse.
- Interpreting travel narratives to contextualize textual references within historical geographies.

Current Progress and Milestones Achieved

- Successful cleaning and pre-processing of Early Modern texts.
- BART and GPT-4o models fine-tuned for OCR correction and thematic analysis.
- Completion of key readings and extraction of structured outputs from rich primary and secondary historical sources.
- Identified and tagged ethically charged language across texts.
- Created maps showing key events in both the world and Virginia.
- Finalized the team's Github here.
- Poster and Artifacts submitted.

Methodology and Approach

The research involves a dual-track methodology:

1. **Historical Analysis**: In-depth review of primary travel narratives, propaganda, and trade documents from the 16th–17th centuries, supported by secondary scholarship. Decided to focus on 1606-1626 where the planning and execution of the massacre happened.
2. **Computational Analysis**: Fine-tuning BART on cleaned OCR text, followed by topic modeling and ethical sentiment detection. Geographical references are extracted and aligned with modern cartographic data using ArcGIS and QGIS.

Shiny apps are being prototyped to allow users to explore the shifting discourse interactively.

Preliminary Results or Findings

- Ethical language around consumption was highly gendered and racialized.
- Trade texts often employed spiritual justifications to mask economic motivations.
- Patterns indicate moral ambivalence surrounding goods obtained via coercive labor.
- Place-name mentions in narratives often reflect colonial ideology and Eurocentric cartographic biases.

Timeline and Schedule

- **July 21**: Poster submission
- **July 26**: Final poster presentation day

Challenges and Obstacles Encountered

- **Textual Complexity**: Extremely large corpora and conflicting historical accounts pose summarization and alignment challenges.
- **Mapping Issues**:
    - Many locations lack definite or standardized coordinates.
    - Temporal ambiguity and shifting borders reduce spatial accuracy.
    - Historical narratives often describe space metaphorically, not literally.

Solutions:

- Use river systems as stable geographic anchors.
- Reconstruct village distribution based on cross-referenced archaeological and textual data.

Resources and Tools Used

- **Models**: BART (fine-tuned), GPT-4o (few-shot model)
- **Mapping**: ArQGIS, RShiny
- **Corpora**: Digitized Early Modern English texts, travelogues, merchant records.

Conclusions

- **Text Cleaning:** The performance scores of both GPT- and BART-cleaned texts reveal that our NER model struggles with the VCR regardless of text cleaning. Notably, the model struggles with identifying LOC entities and the "ambiguity" of Indigenous names (ex. "Powhatan" could be a GPE, PERSON, or NORP). To minimize the need to manually search for locations and their name variations in the VCR, we recommend pre-training an NER model on the VCR.
- **Mapping:** The evolution of key locations as visualized by our map is mostly consistent with the colony's history. For example, settlements most affected by the Massacre of 1622 (ex. Martin's Hundred, Warrascoyack, Henrico) increased the most in mention density post-1622. However, the VCR's introductions, annotations, and titles add noise to our data. For example, Somers Isles (the most common location in titles) is the most mentioned location in 8 out of 18 recorded years and experiences an outsized spike in frequency from 1622-1623. Therefore, we additionally recommend an automated filtering process for non-primary source texts in the VCR.

Final code can be found in our Github