# Mining Historical Testimonials and Visuals to Develop Predictive Models of Genocide Risk

## Introduction

Genocides are among the most monstrous failures of humanity but predicting and preventing them continue to present complex challenges. Past genocides have produced a massive amount of qualitative and visual data, from survivor testimonials to photographs and videos. These sources elicit rich, multidimensional narratives of the events leading up to genocide, and they help to contextualize the sociopolitical and cultural conditions that quantitative datasets often fail to capture.

In this proposal, we propose to mine thousands of testimonials and visuals from the past genocides to catch patterns in this data and form a predictive logistic regression model in SAS. By uniting qualitative and quantitative methods, this study aims to create a tool for policymakers to measure the probability of genocide within particular vulnerable regions.

## Research Objectives

To trace de facto precursors and trajectories based on patterns occurring in testimonial and visual data of historical genocides. Create a logistic regression model to predict genocide risk based on thematic patterns we learned about from the data. Aim to provide actionable insights to international organizations and governments in the field of genocide prevention.

## Conceptual Framework

The study draws on the Genocide Risk Factor Model of Populist Leaders from the proposers' dissertation, extending its use beyond unstructured qualitative and visual data. We apply an adapted Constructed Actor Model of Main Theories of Genocide and the underlying theoretical framework such that leaders, social division, and enabling environments interact dynamically throughout genocide's phases.

## Data Collection

Testimonies: Survivor accounts from archives like the USC Shoah Foundation, Human Rights Watch and of trials.

Visuals: Photographs, videos, sketches showing the events, extracted from NGO reports, media sources, historical records.

## Data Preprocessing

Textual Data: Apply natural language processing (NLP) for sentiment analysis, topic modeling, and theme extraction. Use qualitative coding to identify narratives of early warning signals, including hate speech or who is being targeted.

Visual Data: Perform image classification and mass object classification with computer vision techniques (such as TensorFlow- to isolate recurring images such as gatherings, destruction of property, etc.).

## Methodology

### Phase 1: Descriptive Analysis

Qualitative Analysis: Employ thematic analysis to uncover common patterns in testimonials, that is, propaganda narratives, warning signs, and perpetrator participation. Use qualitative software like NVivo or Atlas to categorize patterns. ti.

Visual Analysis: Use classification to identify hotspots/critical elements (Armed groups/mass displacement).

### Phase 2: Quantitative Modeling

Variable Construction: Make qualitative themes quantifiable as variables (e.g., frequency of hate speech, militarized imagery present)

Logistic Regression Model: Genocide occurrence or non-occurrence.

Methods I factor: Thematic and visual features, and contextual variables (e.g., political regime type, economic instability).

Model Validation: Evaluate by applying the model against genocides prior to the dataset training cut-off date.

### Phase 3: Policy Application

Build a dashboard that will help analyze the risk levels according to model prediction. Work with policymakers to embed findings into early warning systems.

### Ethical Considerations

Data Sensitivity: Use survivor testimonials and visuals, if necessary, judiciously and with permission.

Anonymization: Anonymize personal data to protect survivors.

Bias Mitigation: Consider potential biases in the data collection or analysis.

### Resources and Tools

Textual Data: NLP and text-mining in Python and R Image classification (TensorFlow or PyTorch)

Modeling: SAS for logistic regression and statistical analysis. NVivo or Atlas for qualitative analysis. ti for thematic coding.

### Expected Contributions

Descriptive insights: Qualitative and visual data for a more nuanced understanding of genocide precursors

Predictive Tool: Logistic regression model to predict genocide risk, allowing for data on genocide prevention strategies.

Contribution to Knowledge: Developing the theories of genocide risk to include the qualitative and visual aspects.

**Limitations**

The proposed research faces several limitations. First, the reliance on historical testimonials and visuals introduces challenges in data access, authenticity, and completeness, as some records may be biased, fragmented, or unavailable. Second, the subjectivity inherent in qualitative coding and thematic analysis may lead to inconsistencies, even with advanced software. Third, while predictive logistic regression models provide valuable insights, they are constrained by their reliance on historical patterns, which may fail to capture novel dynamics in future genocides. Furthermore, the integration of findings into policymaking could encounter resistance due to political sensitivities and the practical challenges of operationalizing complex predictive tools. Lastly, ethical considerations, including the respectful use of sensitive data and ensuring survivor privacy, present ongoing challenges that must be addressed throughout the research process.

**Timeline**

**Phase 1 Data Collection and Preprocessing** Months 1- 2

Secure data access and authorization.  Prepare textual and visual inputs.

**Phase 2: Descriptive and Quantitative Analysis** (Months 2-6)

Qualitative coding and thematic analysis.  Fit and validate the logistic regression model.

**Stage 3: Policy Recommendations and Dissemination** (Months 6-9)

Develop policy briefs and an interactive dashboard.  Disseminate findings at academic and policy events.

**Conclusion**

The potential for qualitative and visual data to help improve these models has not yet been exploited. Through descriptive insights that were supplemented with predictive modeling, this research offers a novel approach to identifying and countering genocidal risks. The research will benefit academic scholarship, policymaking and worldwide initiatives to prevent mass atrocities.