README FILE:

Synthetic Dataset for Virtual Team Communication Study

Overview

This synthetic dataset was generated for the study titled "Communication Adaptation in Virtual Teams: An Agent-Based Modeling Approach Using Synthetic Data". The dataset consists of 5000 agents, representing virtual team members, and an additional 500-agent subset used for Agent-Based Modeling (ABM) simulations. The dataset captures key variables related to communication preferences, adaptation patterns, and interaction dynamics over time.

Dataset Files

- synthetic_5000.csv: Full dataset containing 5000 virtual team members with predefined attributes.
- abm_500.csv: Subset of 500 agents used for ABM simulations over 20-time steps.
- variable_description.txt: Detailed descriptions of each variable in the datasets.
- simulation_log.txt: Summary of ABM iterations and recorded adaptations.

Data Structure

Each dataset includes the following variables:

- Agent_ID: Unique identifier for each virtual team member.
- Language_Proficiency: A scale from 1 (low) to 5 (high), representing the agent's ability to communicate in the team's primary language.
- Preferred_Communication_Medium: Initial communication preference categorized as Text, Voice, or Video.
- Interaction_Frequency: Number of weekly communication instances per agent.
- Adaptation_Probability: Likelihood of an agent switching communication mediums over time.
- Medium_Changes: Number of times an agent changed communication mediums during the simulation.
- Final_Communication_Medium: The medium each agent settled on after the adaptation process.

Data Generation Methodology

A probabilistic modeling approach was employed to generate a dataset of 5,000 synthetic virtual team members using a Generative Pre-trained Transformer (GPT)-based framework. Initial communication preferences were assigned based on empirical distributions from organizational communication research. The ABM subset was further processed through 20 simulation time steps, where agents adjusted their communication choices based on:

1. Peer influence (adopting widely used media in their network).
2. Organizational constraints (mandatory video meetings, for example).
3. Individual adaptation tendencies (modeled through stochastic probabilities).

Applications

This dataset is designed to support research in:

- Virtual team dynamics and communication behavior.
- Agent-based simulations of workplace interactions.
- Impact assessment of organizational communication policies.
- Machine learning applications for predicting communication adaptation trends.

Citation

If you use this dataset, please cite:

Haidemariam T. (2025), "Communication Adaptation in Virtual Teams: An Agent-Based Modeling Approach Using Synthetic Data," [Journal name].

Contact

For any inquiries regarding the dataset, please contact:

- Lead Researcher: Tsehaye Haidemariam

  Email: tsehaye.haidemariam@bi.no

  Institutional Affiliation: BI Norwegian Business School

  Centre for Creative Industries (BI: CCI)

  Department of Communication and Culture

  Office address: Nydalsveien 37, 0484 Oslo, Norway

Variable Descriptions

This document provides a detailed description of the variables included in the synthetic datasets: synthetic_5000.csv and abm_500.csv.

1. Agent_ID

- Description: Unique identifier assigned to each virtual team member.
- Type: Integer
- Example Value: 10234

2. Language_Proficiency

- Description: Represents an agent's ability to communicate in the team's primary language, measured on a scale of 1 to 5.
- Type: Integer (1 = Low, 5 = High)
- Example Value: 3

3. Preferred_Communication_Medium

- Description: The initial communication medium selected by an agent before adaptation.
- Type: Categorical (Text, Voice, Video)
- Example Value: Text

4. Interaction_Frequency

- Description: The number of weekly communication interactions an agent has with team members.
- Type: Integer
- Example Value: 12

5. Adaptation_Probability

- Description: Probability of an agent changing their communication medium over time, influenced by peer interactions and organizational constraints.
- Type: Float (0.0 - 1.0)
- Example Value: 0.45

6. Medium_Changes

- Description: Total number of times an agent changed communication mediums throughout the simulation.
- Type: Integer
- Example Value: 2

7. Final_Communication_Medium

- Description: The communication medium an agent used at the end of the adaptation process.
- Type: Categorical (Text, Voice, Video)
- Example Value: Video

Simulation Log

This document provides a summary of the Agent-Based Modeling (ABM) iterations and recorded adaptations over the simulation period.

Simulation Overview

- Total Agents Simulated: 500
- Total Time Steps: 20
- **Initial Communication Distribution in ABM (Time Step 0):**
    - Text: 35.4%
    - Voice: 34.4%

    - Video: 30.2%
- **Final Communication Distribution (Time Step 20):**
    - Text: 42.6 %
    - Voice: 31.2 %
    - Video: 26.2 %

Adaptation Trends

- Percentage of Agents Who Changed Communication Medium: 68%
- Average Number of Adaptations per Agent: 1.7
- Most Common Adaptation: Text → Voice
- Least Common Adaptation: Voice → Text

Organizational Influence

- Mandatory Video Policy Introduced at Time Step 10
    - Led to a 15% increase in video adoption.
    - 8% of agents exhibited reluctance by reducing interaction frequency.

Peer Influence Effects

- Agents in High-Proficiency Groups had a higher adaptation rate (72%) compared to Low-Proficiency Groups (59%).
- Strong network clusters accelerated communication adaptation.

Final Observations

- Communication adaptation follows a gradual process rather than an immediate shift.
- Organizational constraints significantly accelerate media transitions, but peer influence plays a dominant role over time.
- Text-based communication remains prevalent due to its efficiency and accessibility, despite an increase in voice and video usage.