README:

Human–AI Collaboration and the Creativity–Legitimacy Paradox: Synthetic Evidence from Creative Work

Overview:

This repository contains the synthetic stimuli, synthetic respondent ratings, merged analysis dataset, and figures for the paper: Human–AI Collaboration and the Creativity–Legitimacy Paradox: Synthetic Evidence from Creative Work.

This project implements a full-factorial synthetic experiment to examine how configurations of creative agency, algorithmic aesthetics, and infrastructural disclosure shape cultural evaluations of creative advertising work.

- Design: 3 (Authorship / Collaboration mode) × 3 (Aesthetic regime) × 2 (AI disclosure)
- Stimuli: 108 short advertising texts (6 per condition) for fictional brands
- Evaluators: 1,080 synthetic evaluations (10 per stimulus), role-conditioned as: Consumers (50%), Creative directors (25%), Art critics (25%).

Dataset Files:

- Stimulus_plan_108_with_brands.csv
- Completed_stimulus_file.csv
- Ratings_synthetic_respondents.csv
- Merged_stimuli_ratings.csv
- Full_factorial_conditions.csv

Figures:

- legitimacy_interaction_plot.png
- creativity_interaction_plot.png
- originality_interaction_plot.png
- role_moderation_legitimacy.png
- summary_divergence_figure.png
- legitimacy_interaction.png
- legitimacy_coefplot.png
- creativity_coefplot.png
- originality_coefplot.png

Files Included

Core datasets

1. Data/full_factorial_conditions.csv

Full factorial condition table (18 unique conditions) with factor levels.

2. Stimulus_plan_108_with_brands.csv

Stimulus plan expanded to 108 rows (6 per condition), including fictional brand + category assignments.

3. Completed_stimulus_file.csv

Completed stimulus file including stimulus text for each stimulus.

4. Ratings_synthetic_respondents.csv

Synthetic respondent ratings (long format), including role labels and item-level ratings.

5. Merged_stimuli_ratings.csv

Merged analysis-ready dataset (stimuli + ratings). This is the file used for modeling.

Experimental Factors

A) Collaboration mode / authorship framing (3 levels)

- `human_only`
- `ai_only`
- `human_ai_collab`

B) Algorithmic aesthetics (3 levels)

- `mainstream_canon`
- `subcultural_niche`
- `dataset_biased`

C) AI disclosure (2 levels)

- `disclosed`
- `not_disclosed`

Data Dictionary (Key Variables)

Stimuli files (`completed_stimulus_file.csv`)

- `stimulus_id`: Unique identifier for each stimulus (1–108)
- `condition_id`: Condition identifier (maps into the 18-condition table)
- `brand_name`: Fictional brand name
- `product_category`: Fictional product category
- `collab_mode`: Authorship framing (human_only / ai_only / human_ai_collab)
- `aesthetic_regime`: Aesthetic regime (mainstream_canon / subcultural_niche / dataset_biased)
- `disclosure`: AI disclosure condition (disclosed / not_disclosed)
- `stimulus_text`: Full advertising copy (90–120 words), including positioning + tagline + sensory detail

Ratings file (`design1_ratings_synthetic_respondents.csv`)

- `response_id`: Unique evaluation record

- `stimulus_id`: Foreign key to stimulus
- `respondent_id`: Synthetic respondent identifier
- `role`: consumer / creative_director / art_critic
- `creativity`: 1–7
- `originality`: 1–7
- `authenticity`: 1–7
- `cultural_value`: 1–7
- `legitimacy`: 1–7
- `justification_text`: 1–2 sentence rationale (optional qualitative check)

Merged analysis file (`design1_merged_stimuli_ratings.csv`)

- Includes all variables above, allowing one-row-per-evaluation modeling.

Reproducibility Notes

What is synthetic?

Both (a) the creative stimuli and (b) the respondent evaluations are synthetically generated.

All brands, categories, and texts are fictional by design to avoid brand familiarity confounds.

Randomness / determinism

The datasets in `data/` represent a fixed synthetic realization of the design. If you regenerate stimuli or ratings with a different seed, outputs may differ while preserving the same factorial structure.

Suggested analysis approach

Primary analyses in the paper use:

- OLS regressions with stimulus-clustered standard errors
- Interaction tests (collab_mode × disclosure)
- Role moderation models
- Visualization via interaction plots and coefficient plots

How to Use This Repo

Option A: Use the analysis-ready file

If you only want to reproduce models/figures from the current dataset:

1. Load: `data/design1_merged_stimuli_ratings.csv`

2. Estimate models (one per DV: legitimacy, authenticity, creativity, originality, cultural_value)

3. Cluster SEs at `stimulus_id`

Option B: Inspect the generation pipeline artifacts

- Start with `design1_full_factorial_conditions.csv`

- Expand into `design1_stimulus_plan_108_with_brands.csv`
- Review texts in `design1_completed_stimulus_file.csv`
- Review ratings in `design1_ratings_synthetic_respondents.csv`
- Use merged file for modeling

Citation

If you use or adapt this dataset, please cite the associated paper:

- *Author(s). (Year). "Human–AI Collaboration and the Creativity–Legitimacy Paradox: Synthetic Evidence from Creative Work" (Working paper).*

(Replace with the final bibliographic details when available.)

Contact

For questions, issues, or collaboration:

Tsehaye Haidemariam

BI Norwegian Business School

E-Mail: tsehaye.hailemariam@bi.no