RAISE-26

Hosted By Rutgers Bloustein MPI

https://go.rutgers.edu/RAISr E-26
https://go.rutgers.edu/rutgersinfx

Mirror, Mirror on the Wall, Is AI Transforming Us All?

Exploring AI's Impact on Human Behavior

Competition Overview

Background:
The Master of Public Informatics Program at the Rutgers Bloustein School proudly announces RAISE-26, an innovative Informatics, Data Science, and Artificial Intelligence (AI) competition where students are tasked to envision the scope of AI's transformative impact on human society. With the theme "Mirror, Mirror on the Wall, Is AI Transforming Us All?," RAISE-26 challenges participants to tackle one of the most profound questions of our time: How will AI impact human behavior?

Task:
Participants will be provided with a dataset (see note on dataset choice below) containing news headlines. Their task is to examine AI's impact on human behavior and focus on the challenge of determining how AI is shaping the ways humans think, act, and interact. The challenge involves using analytical text analytics and natural language processing (NLP) techniques to extract meaningful insights from these headlines. The goal is to generate innovative data-supported insights that deepen our understanding of the possibilities of an AI-integrated future, promoting informed discussions and strategic decision-making to ensure equitable applications of AI shared by society.

AI use policy:
Teams may use LLMs, multimodal foundation model applications, and tools like Claude, Gemini, ChatGPT, etc. However, the use of any AI tool/s MUST be explicitly declared along with a list of prompts submitted as an appendix. Any use of AI without explicit declaration of specific prompts is strictly prohibited. All references must be cited using APA 6 or higher. Even with fair use of AI, there must be sufficient and dominant application of human intelligence and expertise.
Use of generative AI is thus conditionally permitted, subject to full disclosure and dominant human oversight.

Assistance from any External Help outside of the team:

a. Direct involvement in the task or creation of deliverables is strictly not permitted.
b. However, receiving guidance/seeking advice is acceptable.
c. Faculty advisors to teams are permissible, subject to a. and b. Above.

RAISE-26

Hosted By Rutgers Bloustein MPI

https://go.rutgers.edu/RAISE-26
https://go.rutgers.edu/rutgersinfx

Dataset Description

NOTE: These RAISE-26 datasets are confidential and must not be shared publicly to any forum (e.g., no uploads to GitHub, Kaggle, Hugging Face, or any public forum). Datasets MUST be used only for this competition. After the competition concludes, the RAISE-26 team will publish the dataset publicly (official release). Until then, all participants must treat all the data provided by RAISR-26 as restricted.

- **Dataset A:** Full news dataset containing **10,500 rows** (news headlines) across **11 columns**.
- **Dataset B:** A **3,500-row subset** of Dataset A with the **same 11 columns**; it contains **no additional/unique records** beyond Dataset A.

The dataset consists of a curated collection of **news headlines related to Artificial Intelligence (AI)**. Each record represents one AI-related news headline, enriched with linguistic and temporal features to support analyses of how AI is portrayed in the media and how it relates to human behavior, emotion, and societal context.

1. **date**: The publication date of the news article in YYYY-MM-DD format.
2. **title**: The headline of the news article.
3. **source**: The publisher or platform where the news article was published.
4. **number_of_characters_title**: The total number of characters in the headline, including spaces and punctuation.
5. **number_of_words_title**: The total number of words in the headline.
6. **day_of_week**: The day of the week corresponding to the publication date (e.g., Monday, Tuesday).
7. **month**: The month in which the news article was published (e.g., January, February).
8. **year**: The year in which the news article was published.
9. **quarter**: The quarter of the year when the news article was published (e.g., 1 for Jan-Mar, 2 for Apr-Jun).
10. **is_weekend**: A boolean-like value (True or False) indicating whether the news article was published on a weekend.
11. **classes_str**: A semicolon-separated list of one or more conceptual categories (or "classes") identified in the headline based on keyword matching.

Each headline may belong to multiple classes if it contains keywords from more than one category.

Classes

Each class represents a thematic dimension describing how AI is framed with respect to human behavior, emotion, and society:
(NOTE: Classes are directional and expected to lead to richer analyses. This classification is done using keyword-based approximations. Headlines and classes have not been manually validated, so some entries may be misclassified or only partially aligned with the intended theme. Therefore, the "classes_str" column should be used carefully, especially for detailed analyses. Teams may feel free to agree/disagree with, refine, or omit the use of classes.)

1. **Sentiment (Positive / Negative Feelings)**: Headlines expressing emotional valence such as optimism, trust, fear, anxiety, disruption, or bias.
2. **Emotion, Motivation & Well-being**: Mentions of mental states, stress, motivation, confidence, mindfulness, or emotional regulation.
3. **Human Roles**: References to people and social identities such as workers, students, parents, doctors, or creators.
4. **Routine, Lifestyle & Behavior**: Mentions of daily habits, productivity, sleep, diet, fitness, or consumer activities.
5. **Cognitive & Decision-Making**: Focus on thinking, reasoning, judgment, bias, or problem-solving processes.
6. **Creativity, Expression & Identity**: Headlines linking AI with creativity, art, design, authenticity, or personal identity.
7. **Social Interaction & Relationships**: Discussions of communication, empathy, social media, or human–AI interaction.
8. **Work, Jobs & Economy**: Coverage related to employment, automation, productivity, layoffs, or the broader economy.
9. **Learning, Knowledge & Education**: Topics involving education, learning, curiosity, or academic performance.
10. **Health, Safety & Risk**: Mentions of physical or mental health, safety, medical care, or risk.
11. **Society, Ethics & Culture**: Broader societal issues such as privacy, fairness, governance, rights, or cultural norms.
12. **Technology & Interaction**: Headlines about digital technologies, devices, social platforms, or virtual/augmented environments.

https://go.rutgers.edu/RAISE-26
https://go.rutgers.edu/rutgersinfx

Submission Guidelines

1. **Core Requirement:** Submissions may be validated by the judges, so please test your code before submitting. All files must run independently and reliably reproduce the same outputs and results that you report in your notebook and slides.
2. **Dataset Requirements:**
   - 2.1. **Mandatory dataset usage:** Teams must use Dataset A or Dataset B. Submissions cannot be based only on Dataset C (Synthetic). The available datasets are A (full news dataset), B (subset of A), and C (synthetic add-on). Using C is optional and it will be released in January.
   - 2.2. **Dataset sizes and purpose:** Dataset A contains 10,500 rows, and Dataset B contains 3,500 rows with the same schema. Dataset B is provided for teams facing technical constraints (e.g., limited compute/RAM such as <16–32 GB). This is generally not a limitation for teams using Google Colab.
   - 2.3. **Fairness in judging**: Judging will not differentiate submissions based on whether teams used Dataset A or Dataset B. Teams must clearly state in their PowerPoint which dataset(s) they used.
3. **Reproducibility Rules for LLMs:**
   - 3.1. Non–open-weights LLMs are rejected due to reproducibility concerns; examples of **not allowed** closed/hosted-only models include OpenAI models (gpt-4o, gpt-4.1, gpt-3.5-turbo, chatgpt-4o-latest), Anthropic models (claude-3-5-sonnet, claude-sonnet-4-5), Google Gemini API-hosted models (gemini-2.5-pro, gemini-1.5-pro-002), Cohere API-hosted models (command-r-plus-04-2024, command-r-plus), Perplexity API models (pplx-70b-online), and xAI's Grok-2 (hosted/proprietary).
   - 3.2. If an LLM is used, it must be open-weights, or downloadable, the specific version must be clearly stated (exact repo + version tag/revision), and it must be run within the notebook/script environment (weights downloaded/loaded there) or invoked via the Hugging Face Inference API only if the model is open-weights, version-pinned, and still downloadable; examples of allowed open-weights models include mistralai/Mistral-7B-Instruct-v0.3, mistralai/Mixtral-8x7B-Instruct-v0.1, Qwen/Qwen2.5-7B-Instruct, meta-llama/Llama-3.1-8B-Instruct, google/gemma-2-9b-it, google/gemma-2-27b-it, microsoft/Phi-3.5-mini-instruct, allenai/OLMo-7B-Instruct.
4. **Execution & Environment Expectations:**
   - 4.1. **Notebook format and "one-click run" (Mandatory)**
     - **4.1.1.** Preferred tool for RAISE-26: Google Colab .ipynb.
     - 4.1.2. Submissions must be tutorial-style with notes and must support "one-click run" end-to-end via Colab → "Run all."

4.1.3. Test your notebook before final submission.

4.1.4. The notebook must run without judges adding, editing, or changing any details, including tokens/keys.

4.2. Authentication, tokens, and access (If applicable)

4.2.1. Token-free execution (Strongly recommended): Where viable, use public models/datasets and load them without authentication so the notebook runs fully without tokens.

4.2.2. If authentication is unavoidable: If authentication is required (e.g., gated access or rate limits), include a clearly labeled "Authentication" cell with setup instructions. Any required token(s) must be provided in the submission so judges can run the notebook seamlessly (no manual token entry/editing required).

4.2.3. Token security and responsibility: If tokens are included, apply appropriate safeguards (e.g., spending/dollar limits where supported). RAISE-26 is not responsible for any token usage under any circumstances. After the conclusion of RAISE-26 (April 17, 2025), participants are responsible for promptly removing or deactivating all API keys/tokens associated with RAISE-26 Colab submissions from any platforms used (e.g., Hugging Face) to prevent future unauthorized use.

4.3. Agentic workflows (If applicable)

4.3.1. Single-notebook requirement: If your project is an agentic workflow, the entire workflow must run in ONE Colab notebook and must execute end-to-end via "Run all" with no judge modifications (including tokens).

4.3.2. Permitted agentic frameworks: Agentic frameworks (e.g., LangChain/LangGraph, LlamaIndex, CrewAI etc) are allowed only if they run end-to-end in the single Colab notebook and use open-weights models and provided they comply with the LLM reproducibility and authentication rules in Sections 4.1-4.2 and the workflow runs end-to-end in a single notebook.

5. Deliverables (Both are mandatory) Each team must submit:

5.1. Full Script (Google Colab Python Notebook – .ipynb): One .ipynb file is preferred, but multiple files are allowed. The entire end-to-end workflow must run from a single Google Colab notebook via **"Run all"** (single-click). Supporting files (e.g., helper scripts/modules) may be included, but the notebook must be the main integration point and must produce and reproduce all results.

5.2. PowerPoint Slides (Report of Findings): Submit slides in PowerPoint format only (.ppt/.pptx) with no limit on the number of slides. The deck is expected to include outputs generated from your code and, where needed, screenshots of relevant code sections, and must provide a detailed report of all findings. Do not submit PDFs or any other slide formats.

6. Permitted Tools and Languages Only the following are permitted:

6.1. Python using Google Colab (and associated libraries) [Preferred for RAISE-26]
6.2. If using LLMs or AI agents, they must be run according to the code/tools in Submission Guidelines #3, #4, or #5.1 (not external chat tools)
6.3. R limited to .R or .RMD files only (and associated libraries/packages)

No other tools or software are permitted.

Please direct all personal team-level questions to: informatics@ejb.rutgers.edu

For updates and networking, please follow us on:
https://www.linkedin.com/company/rutgers masters-in-public-informatics/

Thank you for joining us in this stimulating journey to discover the truths embedded in AI narratives in the news media. Let's collectively shape a future where AI, data science and public-centered informatics empower societal advancements!

Hosted By Rutgers Bloustein MPI
https://bloustein.rutgers.edu/graduate/public-informatics/