Intelligence Advanced Research Projects Activity

I A R P A

Creating Advantage through Research and Technology

HIATUS

HUMAN INTERPRETABLE ATTRIBUTION OF TEXT USING UNDERLYING STRUCTURE

INTELLIGENCE VALUE

The HIATUS program aims to develop novel human-useable systems for attributing authorship and protecting author privacy. Authorship attribution capabilities address many Intelligence Community (IC) needs, including combating sophisticated malicious information campaigns online and identifying counterintelligence risks. Authorship privacy capabilities protect authors whose writing, if attributed, could place them in danger.

Humans and machines produce vast amounts of text content every day. Text contains linguistic features that can reveal author identity. To support and protect the IC mission, the HIATUS program's objective is to develop multi-language-capable tools to attribute authorship and protect author privacy. These tools must implement novel explainable Artificial Intelligence techniques to provide trustworthy and verifiable results to human users regardless of author background or document genre, topic, and length.

The HIATUS program casts authorship attribution and privacy as different aspects of the same underlying challenge: understanding author-level linguistic variation by elucidating stable identifiers of individual authors across diverse types of text. The program places Performers' authorship attribution and privacy systems in competition with one another. Performer teams compete to generate higher fidelity representations between individual authors' unique linguistic fingerprints.

Performer systems are submitted to the HIATUS Testing & Evaluation (T&E) teams for blind evaluation against opponent team systems on a sequestered dataset comprising multilingual documents representing diverse text and author characteristics. Attribution systems are evaluated on ability to match items by the same author in large collections, while privacy systems are evaluated on ability to thwart attribution systems. System explainability will be evaluated using a protocol developed by Performers, T&E teams and Government partners in the beginning of the program.

The HIATUS program began in late 2022 and has a duration of 45 months. The program comprises three phases, including an initial 21-month long phase that focuses on English and two subsequent 12-month long phases that include multiple foreign languages. The program is pursuing multiple venues for early transition of technology to Government partners.

PRIME PERFORMERS

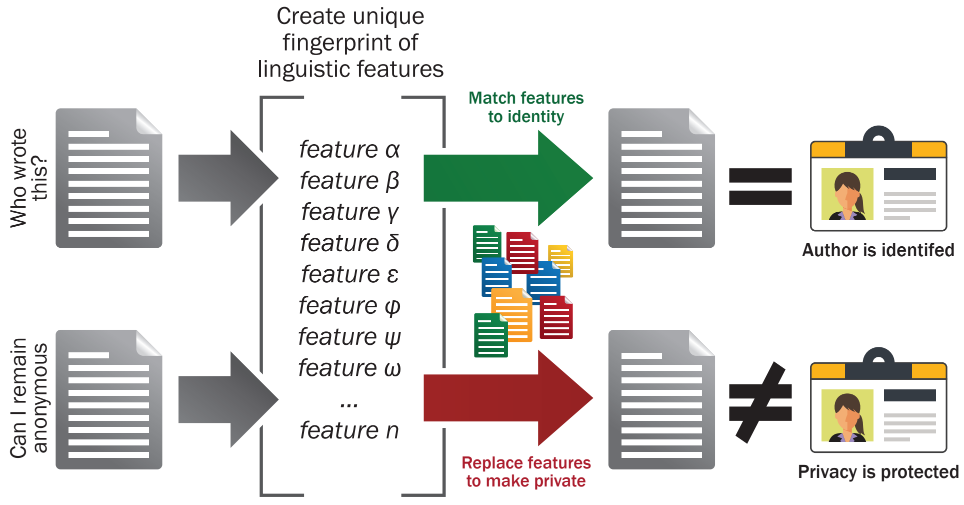
- Charles River Analytics, Inc.
- Leidos, Inc.
- Raytheon BBN
- SRI International
- University of Pennsylvania
- University of Southern California

TESTING AND EVALUATION PARTNERS

- Lawrence Livermore National Labs (U.S. Department of Energy)
- Pacific Northwest National Labs (U.S. Department of Energy)
- University of Maryland's Applied Research Laboratory for Intelligence and Security

KEYWORDS

- Adversarial Machine Learning
- Authorship Attribution
- Explainable AI
- Forensic Linguistics
- Human Language Technology
- Privacy



The HIATUS vision: A combined authorship attribution and privacy system that can be trusted and audited by human operators

PROGRAM MANAGER

Timothy McKinnon, Ph.D.

Phone: (301) 243-2084

timothy.mickinnon@iarpa.gov

www.iarpa.gov

@IARPAnews

Latest Revision: 03/08/2024