James "Jamie" A. Scharf

1441 U St. APT 405 | Washington, DC 20009 | Mobile: (914)-522-2930 | scharfjamie@gmail.com |
https://github.com/JamesScharf

BOOK

Giovanna Maria Dora Dore, Arya D. McCarthy, & **James A. Scharf**. "A Free Press, If You Can Keep It—What Natural Language Processing Reveals about Freedom of the Press in Hong Kong." Springer Briefs in Political Science. Springer Nature (2023). Book.

ARTICLES, CONFERENCE PROCEEDINGS, AND PAPERS

Adam Sheingate, **James A. Scharf**, & Conner Delahanty. (2022). "Digital Advertising in U.S. Federal Elections, 2004-2020." *Journal of Quantitative Description: Digital Media*. https://doi.org/10.51685/jqd.2022.026. Paper.

Arya D. McCarthy, **James A. Scharf**, and Giovanna Maria Dora Dore. "A Mixed-Methods Analysis of Western and Hong Kong–based Reporting on the 2019–2020 Protests." *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Conference on Empirical Methods in Natural Language Processing (EMNLP 2021). Paper. Presentation slides.

James A. Scharf, Arya D. McCarthy, and Giovanna Maria *Dora Dore.* "Characterizing News Portrayal of Civil Unrest in Hong Kong, 1998–2020." *Proceedings of Challenges and Applications of Automated Extraction of Socio-political Events from Text (CASE)*. The Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021). Paper. Presentation recording.

James A. Scharf, Giovanna Maria Dora Dore, and Arya McCarthy. "Evolution and Bias: News Portrayal of Civil Unrest in Hong Kong, 1998-2020." 4th Politics and Computational Social Science Conference (PaCSS 2021). (Peer-reviewed paper).

Arya D. McCarthy, **James A. Scharf**, and Giovanna Maria Dora Dore. Characterizing News Portrayal of Civil Unrest in Hong Kong 2019–2020. (Unpublished)

CONFERENCE PRESENTATIONS AND POSTERS

Arya D. McCarthy, **James A. Scharf**, and Giovanna Maria Dora Dore. "Two Lenses on the Fragrant Harbour: Differences in Western and Hong Kong-based Reporting on the 2019-2020 Protests." *Proceedings of ACL Special Interest Group on Language Technologies for the Socio-Economic Sciences and Humanities* (SIGHUM 2021). Poster.

Arya D. McCarthy, **James A. Scharf,** and Giovanna Maria Dora Dore. Measuring Differences in News Portrayal of the 2019–2020 Hong Kong Protests. New Directions in Analyzing Text as Data (TADA) 2021.

David Yarowsky, Arya D. McCarthy, Garrett Nicolai, Winston Wu, Aaron Mueller, Dylan Lewis, Yingqi Ding, Abhinav Nigam, Emre Ozgu, Debanik Purkayastha, **James A. Scharf** and Kenneth Zheng. A 1000-language Collaborative Universal Dictionary and Universal Translator. UNESCO Language Technologies for All (LT4All) 2019.

EMPLOYMENT & RESEARCH EXPERIENCE

The Johns Hopkins University Applied Physics Lab (APL), Research & Exploratory Development Division (REDD), Artificial Intelligence Group (RQC) – Laurel, MD
July 11 2022 – Present   *Human Language Technology Researcher*
- Principle Investigator of Intelligent Systems Center's first foray into fine-tuning of an LLM with specialized compute
- Implementing complex deep neural networks in PyTorch for tasks such as seq2seq
- Designing and implementing agent-based (e.g., ReAct) LLM systems; e.g., War Room's CoAnalyst, an autonomous, multi-agent LLM-powered system for intelligence
- Regular presenter to high-profile internal and external audiences such as the Office of the Secretary of Defense R&E Executive/Chief Strategy Officer; National Security Council members
- Contributing to research proposals with several government sponsors
- Built zero-shot attribute recognition systems with multimodal models for domain-specific maritime data

Joint MISO WebOps Center (JMWC), U.S. Special Operations Command – Tampa, FL
Summer 2021   *Telesis Civic Digital Fellow, Data Science*
- Created the Named Entity Recognition (NER) and Text Classification machine learning backend for the in-house annotation platform with the Huggingface Transformers, SpaCy, Scikit-Learn, and FlairNLP Python libraries
- Personally designed and implemented a two-path model pipeline for training customized neural and non-neural models based on finetuned BERT-based transformers and conditional random field models
- Added the capability to bootstrap statistical NER, text classification models in a multilingual setting
- Confirmed that trained models could be redeployed and re-trained/finetuned for future objectives

Department of Computer Science, Johns Hopkins – Baltimore, MD
Spring 2021 & Spring 2022   *Head Course Assistant for Information Retrieval and Web Agents*
- Assistant and grader for a course in machine learning, information extraction and web scraping
- Helped students with projects in sense disambiguation, word embeddings and vector models of information retrieval

Department of East Asian Studies, Johns Hopkins – Baltimore, MD
Summer 2020 – Fall 2021   *Computational Research Assistant*

- Built probabilistic models of 4000 news articles with the Causalnex and Scikit-Learn libraries
- Generated plots with Seaborn and Matplotlib to compare news coverage in Hong Kong by year

Center for Language and Speech Processing (CLSP), Johns Hopkins – Baltimore,
Summer 2020   *Intern for Dr. David Yarowsky's Low-Resource Languages Lab (LoReLab)*
- Utilized Facebook's Fairseq Sequence-to-Sequence (Seq2Seq) library for G2P tasks
- Scraped and parsed foreign language webpages into structured data with Pandas, Selenium and BeautifulSoup
- Attended the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)

Department of Political Science, Johns Hopkins University – Baltimore, MD
September 2018 – October 2019   *Research Assistant for Prof. Adam Sheingate*
- Used Python and R to analyze F.E.C. data related to elections from 2010 to 2016
- Built a natural language classifier using the Scikit-learn machine learning framework, pattern matching and query-expansion

ODN (formerly Open Data Nation) – New York, NY
Summer 2019   *Data Science Intern*
- Programmed statistical analyses of roadway and car crash data with Pandas and Scikit-learn
- Cleaned and prepared geospatial datasets for machine learning models

EDUCATION

The Johns Hopkins University – Baltimore, MD   *May 2022*
- Master of Science in Engineering, Computer Science
- Advisor: David Yarowsky at the Center for Language and Speech Processing
- **Master's Project:** *Morphological Annotation of Unseen Lexica by Zero-Shot Generalization of Monolingual Analyzers and Bilingual Morphological Embeddings*. Paper.

The Johns Hopkins University – Baltimore, MD   *May 2021*
- Bachelor of Arts, Political Science; Minor, Computer Science
- Coursework in natural language processing, artificial intelligence, machine translation, and parallel programming
- Advisors and Mentors: Giovanna Maria Dora Dore, Adam Sheingate, David Yarowsky
- **GPA:** 3.56 (General Honors)

MISC. SKILLS & QUALIFICATIONS

Clearance: Interim Top-Secret
Programming Languages: Python, C/C++, Java, Go, JavaScript, HTML, Bash/Unix
Toolkits: PyTorch, Huggingface Transformers, Scikit-Learn, SpaCy, Pandas, Dask, Matplotlib