Patrick **Butler**

SENIOR RESEARCH ASSOCIATE — AI/ML ENTHUSIAST

1366 G St SE, Washington, DC. 20003

(540) 585-4287 | pjbutler@gmail.com | people.cs.vt.edu/pabutler | pbutler | pjbutler | Patrick Butler

Education

May 2014	**PhD in Computer Science**, Virginia Tech
May 2014	**MS in Computer Science**, Virginia Tech
Dec 2005	**BS in Physics**, Virginia Tech
Dec 2005	**BS in Computer Science**, Virginia Tech

Professional Experience

Virginia Tech   *Arlington, VA*

SENIOR RESEARCH ASSOCIATE   *January 2015 - Present*

- Designed a Large Language Model (LLM) pipeline and Retrieval Augmented Generation (RAG) architecture to query open source documents for answering questions related to climate change.
- Used open source data (e.g., Twitter, blogs, news, weather, and other sources) to predict real-world population level events such as civil unrest and rare disease outbreaks in Latin America, making >30,000 predictions in 4 years that were independently validated as part of a $22M IARPA Project. Later, led the team that transitioned this project into a production environment in the government.
- Modeled and predicted COVID-19 and influenza like illness outbreaks during pandemics, producing results which informed public policy.
- Modeled migration patterns on the Southwest border to help inform resource placement during migration booms.
- Researched novel methods of nuclear proliferation detection based on open source news data leveraging word embeddings to detect the semantic shift of words in the source material surrounding points of interest.
- Managed teams of students and organized research for medium sized software oriented research projects.
- Led a research team which designed novel machine learning based information extraction techniques to extract and classify information from scholarly sources. The resulting data allowed government policy makers to make decisions regarding the funding of new research initiatives.

ADJUNCT PROFESSOR IN CS DEPARTMENT

- Taught six semesters of CS5644: Machine Learning & Big Data — Covered basic machine learning techniques using scikit-learn and big data related technologies such as Redis, Elasticsearch, and MapReduce.
- Taught six semesters of CS5664: Social Media Analytics — Covered how to collect, process, and analyze social network related data.
- Each class comprised of around 100 students per semester and required managing a team of three teaching assistants.

US Army Corps of Engineers   *Alexandria, VA*

RESEARCH SCIENTIST   *September 2014 - January 2015*

- Researched methods for geocoding sparsely geotagged data.
- Researched methods for generating analyst-directed stories from a corpus of documents.

Self-Employed   *Blacksburg, VA*

RESEARCH CONSULTANT   *May 2014 - Sept 2014*

- Supported EMBERS project at Virginia Tech and continued research for predicting civil unrest predictions.

Virginia Tech   *Blacksburg, VA*

GRADUATE RESEARCH ASSISTANT   *July 2007 - May 2014*

- Developed algorithms and architecture for using social networks to predict civil unrest and influenza like illnesses.
- Developed algorithms for mining electronic medical records.
- Designed a tool for graphing and exploring social networks, inferred from one or more inboxes.
- Researched and designed new methods for mining temporal data streams especially for use in bioinformatics.

Professional Organizations

- Association for Computing Machinery
- Upsilon Pi Epsilon

Papers (Selected)

Evaluation of FluSight influenza forecasting in the 2021–22 and 2022–23 seasons with a new target laboratory-confirmed influenza hospitalizations

S. M. MATHIS, ET AL

Nature Communications 15.1 (July 2024) p. 6289. 2024

Machine Learning Modeling Pipeline for Extracting Nuclear Proliferation Events of Interest from Open Data Sources (U)

T. L. Danielson, B. Mayer, N. Muralidhar, J. Miller, H. Dogan, N. Self, P. Butler, F. Liu

Tech. rep., 2022

Lessons from Deep Learning applied to Scholarly Information Extraction: What Works, What Doesn't, and Future Directions

R. B. Yousuf, S. Biswas, K. K. Kaushal, J. Dunham, R. Gelles, S. Muthiah, N. Self, P. Butler, N. Ramakrishnan

ACM KDD Workshop on Data-driven Science of Science, 2022

Using AntiPatterns to avoid MLOps Mistakes

N. Muralidhar, S. Muthiah, P. Butler, M. Jain, Y. Yu, K. Burne, W. Li, D. Jones, P. Arunachalam, H. McCormick

arXiv preprint arXiv:2107.00079 (2021). 2021

Detecting Media Self-Censorship without Explicit Training Data

R. Tao, B. Zhou, F. Chen, D. Mares, P. Butler, N. Ramakrishnan, R. Kennedy

Proceedings of the 2020 SIAM International Conference on Data Mining, 2020

Syndromic surveillance of Flu on Twitter using weakly supervised temporal topic models

L. Chen, K. T. Hossain, P. Butler, N. Ramakrishnan, B. A. Prakash

Data mining and knowledge discovery 30.3 (2016) pp. 681–710. Springer US, 2016

Seeing the forest for the trees: new approaches to forecasting cascades

S. Krishnan, P. Butler, R. Tandon, J. Leskovec, N. Ramakrishnan

Proceedings of the 8th ACM Conference on Web Science, 2016

EMBERS at 4 years: Experiences operating an open source indicators forecasting system

S. Muthiah, P. Butler, R. P. Khandpur, P. Saraf, N. Self, A. Rozovskaya, L. Zhao, J. Cadena, C.-T. Lu, A. Vullikanti

Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016

Monitoring disease trends using hospital traffic data from high resolution satellite imagery: a feasibility study

E. O. Nsoesie, P. Butler, N. Ramakrishnan, S. R. Mekaru, J. S. Brownstein

Scientific reports 5.1 (2015) pp. 1–8. Nature Publishing Group, 2015

Forecasting significant societal events using the embers streaming predictive analytics system

A. Doyle, G. Katz, K. Summers, C. Ackermann, I. Zavorin, Z. Lim, S. Muthiah, P. Butler, N. Self, L. Zhao

Big data 2.4 (2014) pp. 185–195. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, 2014

'Beating the news' with EMBERS: forecasting civil unrest using open source indicators

N. Ramakrishnan, P. Butler, S. Muthiah, N. Self, R. Khandpur, P. Saraf, W. Wang, J. Cadena, A. Vullikanti, G. Korkmaz

Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, 2014

The Deshredder: A Visual Analytic Approach to Reconstructing Shredded Documents

P. Butler, P. Chakraborty, N. Ramakrishan

2012 IEEE Conference on Visual Analytics Science and Technology (VAST), 2012

Quantitatively analyzing stealthy communication channels

P. Butler, K. Xu, D. D. Yao

International Conference on Applied Cryptography and Network Security, 2011