Charles Alba

alba@wustl.edu | sites.wustl.edu/alba | Google Scholar | cja5553 | charles-alba-24bb63186
| 05 Brown Hall, St. Louis, MO 63105 | cja5553

Education

PhD in Computational & Data Sciences *St Louis City, MO, USA*
Washington University in St Louis 2022 - present
- GPA: 4.0/4.0

MS in Behavioral & Data Science *(Distinction)* *Coventry, England, UK*
University of Warwick 2021 - 2022
- Thesis: "The role of default-displayed reviews in anchoring online reviewers: text evidence from STEAM video-games"

BS in Data Science *University Park, PA, USA*
The Pennsylvania State University 2018 – 2021
- Minor in Statistics
- GPA: 3.83/4.0
- Mu Sigma Rho Honors

Research Interests

Computational / Methodological expertise
- Natural language processing (NLP), large language models (LLMs) & text mining *[primary]*
- Artificial intelligence & deep learning
- Health & social data sciences

Domain expertise
- Public health & healthcare, with a focus on food assistance policies *[primary]*
- Social determinants of health
- Behavioral sciences

Skills

Technical skills *(analytical)*	Machine learning & AI, deep learning, applied natural language processing (NLP) & text mining, spatial mapping & analytics, social media mining, statistical models & analysis
Technical Skills *(Data Extraction)*	Data wrangling, web scrapping & APIs
Coding skills	Python, R, SQL, NoSQL languages, LaTeX, QGIS
Languages	English (Native), Chinese (proficient)

Academic Research Experience

Danforth Scholar & Graduate Research Assistant *St Louis City, MO, USA*

Division of Computational & Data Science, Washington University Aug 2022 - present

- Selected research projects:
  - Applying large language models to improve Supplemental Nutrition Assistance Program (SNAP) policies [dissertation research]
    - SNAP is a food assistance initiative that aims to reduce food insecurity in America, which predominantly affects socio-economically vulnerable communities. The goal of my dissertation research is to develop large language model (LLM) tools to empower both recipients and policy-makers of SNAP to make informed decisions through textual data.
    - These LLM tools include (1) sentiment classifiers that could understand attitudes, opinions, and emotions through social texts, (2) concise LLM-powered topic models that could perform topic extraction on the most prominent underlying factors resulting inequitable SNAP outcomes through vast text-based literature, and (3) grounded Q&A chatbots to help recipients navigate the complexities of SNAP.
  - Applying large language models (LLMs) on surgical text to predict perioperative care from patients
    - We used ∼85k clinical notes from the BJC Healthcare system to build predictive LLMs. These models aim to predict surgical complications and outcomes of patients namely mortality, Deep vein thrombosis (DVT), pulmonary embolism (PE), pneumonia, acute knee injury (AKI) & Delirium.
    - We applied state-of-the-art pre-trained LLMs, such as `BioGPT`, `ClinicalBERT`, and `bioClinicalBERT`. Additionally, we experimented with various novel fine-tuning strategies on these pre-trained LLMs, including semi-supervision and a foundation model comprising of multi-task learning (MTL).
    - These strategies resulted in improvements towards prediction performance compared to using pre-trained or traditional fine-tuned models. Specifically, foundational MTL yielded improvements of up to 3% for AUROC and 2.6% for AUPRC.
    - Currently finalizing the results and preparing manuscript for submission.
  - Using mobile phone data to assess disparities in unhealthy food consumption during COVID-19
    - Using county-level longitudinal data on visits to unhealthy food outlets from ∼80k points-of-interests, alongside New York Times' COVID-19 data and socio-economic data from the American Community Survey, we analyzed COVID-19's impact on changes in socio-economic disparities associated with unhealthy food reliance.
    - Results showed convenience store reliance increased amongst college-educated and Hispanic demographics, whereas fast-food reliance increased from the elderly but decreased amongst Hispanic demographics.
    - Paper published in Health Data Science journal.

Graduate Student Assistant *Coventry, England, UK*

Dept of Psychology, University of Warwick Mar 2022 - Aug 2022

- Project: Applying text-mining tools to examining cognitive biases in online reviews
- Advisors: Drs. Lukasz Walasek & Mikhail Spektor
- Project Description:
  - We hypothesized that salience and order effects of online reviews influence users writing new reviews.
  - To test our hypothesis, we scraped ∼1.1 million video game reviews from STEAM's review sorting algorithm to identify reviews that would have been shown to each user when they were writing their review. We then used FastText embeddings and cosine-similarity matrices for similarity assessment.
  - ANOVA results revealed that reviewers imitate the most helpful reviews written by others, especially those that are visually salient. Additionally, the default sorting and display format of reviews on online platforms will have a pronounced effect on the style and content of new reviews.
  - Paper in R&R at the Decisions journal (special issue of Machine Learning, AI & Judgement Decision Making).

Undergraduate Research Assistant *University Park, PA, USA*

Dept of Recreation, Parks & Tourism Management, Penn State University May 2021 – Aug 2021

- Project: Using mobile phone data to assess recreational inequity in national park visitations
- Advisor: Dr. Bing Pan
- Project Description:
  - Analyze the impact of COVID-19 on national park visitations using mobile phone data. The data was selected from ∼40 million points-of-interests (POIs) using shapefiles provided by the National Park Service (NPS).
  - Together with demographic data from the US census, this data was subsequently analyzed using the 'gravity-model' panel data analysis.
  - Our findings revealed a significant increase in visits from communities situated within a 452km radius of a National Park. In contrast, there was a noticeable decline in visitations from non-white and Native American communities, especially for communities located more than 317km and 482km away from a national park, respectively.
  - Paper published in Scientific reports

Undergraduate Research Program *University Park, PA, USA*

Student Engagement Network, Penn State University Jan 2021 – May 2021

- Project: Using quantifiable behavioral traits to predict the spread of COVID-19

Publications

Published Journal Articles

[1] **C Alba**, R An. "Using Mobile Phone Data to Assess Socio-Economic Disparities in Unhealthy Food Reliance during the COVID-19 Pandemic," *Health Data Science [Vol 3, no 101]*, 2023.
10.34133/hds.0101 / Code & data

[2] **C Alba**, B Pan, J Yin, W Rice, P Mitra, M Lin, Y Liang. "COVID-19's impact on visitation behavior to US national parks from communities of color: Evidence from mobile phone data," *Scientific Reports [vol 12, no 13998]*, 2022.
10.1038/s41598-022-16330-z / Code & data / IF: 4.99

[3] **C Alba**, M Mittal. "Sociocultural behavioral traits in modelling the prediction of COVID-19 infection rates," *Journal of Humanities and Applied Social Sciences [vol 3, issue 5]*, 2021.
10.1108/JHASS-07-2021-0128 / Code & data

Articles Under Review

[1] **C Alba**, L Walasek, M Spektor. "Attention-driven imitation in consumer reviews," Revise & Resubmited at *Decision [Special issue on Machine Learning, AI & Judgement Decision-Making]*, 2023.
Code & data

Working Papers

[1] **C Alba**, V Abbasian. 'Applications of Social Media Mining to Uncover Shortfalls in Social Welfare Services to Orphaned Children", 2023.
Code & data

[2] B Xue*, **C Alba***, J Abraham, T Kannampallil, C King, M Avidan, C Lu. "Prescribing Large Language Models for Perioperative Care: What's The Right Dose for Pre-trained Models?", 2023.
Code & data / *Co-first authors

Research Grants / Funding

Internal Grants / Funding

Library Fund, University of Warwick — *£1740*
Project: COVID-19's impact on racial equity to park visitation with mobile phone data — July 2022
- Role: Lead Author (∼PI)

Remote Innovation Grant, Penn State Univ. (Grant No 012599) — *US$1000*
Project: Using quantifiable behavioral traits to predict COVID19 infection rates — Jan 2021 – May 2021
- Role: Student Research Project (∼Student PI)

Awards and Honors

2022-2027	**Scholarship**: "Danforth Scholarship", *Washington University in St Louis*
2022	**Honor**: Distinction (dissertation & degree), *University of Warwick*
2018-2021	**Honor**: "7 ×Deans Lists", *Penn State University*

Teaching

Teaching Assistant — *Washington Univ in St Louis*
MPH 5139 - Applied Machine Learning Using Health Data — Spring 2024

Undergraduate Teaching Assistant — *Dept of Statistics, Penn State*
Stat200 - Elementary Statistics — Fall 2019

Lecturer assistant *Dept of Statistics, Penn State*
Stat200 - Elementary Statistics Spring 2018