

# Kambiz\_Tavabi

Seattle, WA | ktavabi@gmail.com | GitHub | LinkedIn |

---

## PROFESSIONAL SUMMARY

---

Senior Data Scientist and NIH-recognized researcher with 15+ years deploying ML solutions in high-stakes environments from healthcare to corrections. I help organizations make critical decisions faster through production NLP systems and privacy-compliant analytics, with expertise in translational clinical research that led to breakthrough electrophysiological biomarkers for autism spectrum disorders. At Washington DOC, I built ML pipelines that improved incident classification accuracy 30-fold while managing complex multi-table ETL across restricted datasets. My expertise spans end-to-end model development, A/B testing frameworks, and leading cross-functional teams to deliver executive-grade insights under strict regulatory constraints.

## CORE COMPETENCIES

---

Population Health & Healthcare Analytics	Clinical Predictive Modeling (AUC, ROC)
Healthcare Claims Data (APCD, EHR)	Causal Inference & Statistical Modeling
Production ML & Decision Support Systems	Python · pandas · scikit-learn · SQL
A/B Experiment Design & Evaluation	Cross-functional Team Leadership
Speech & Language Processing	Time-Series & Signal Processing

## WORK EXPERIENCE

---

Washington State Department of Corrections

May 2023 – Present

Manager, Business Intelligence & Operations Surveillance (BIOS) – Tumwater, WA

- Deployed production-grade NLP classifier automating incident categorization across 110,928 records and 73 binary labels. Achieved 75.6% micro-F1 and 88.3% precision – a 30-fold improvement over baseline – with ongoing model optimization increasing F1 scores by 5 points through support vector experimentation. Built end-to-end scikit-learn pipeline with TF-IDF features, validated on 27,700 held-out records using k-fold cross-validation with Jaccard scoring, and established human-in-the-loop quality controls for continuous refinement.
- Architected SQL ETL pipelines (T-SQL, Oracle) supporting Power BI dashboards and analytics. Automated gateway refreshes, eliminating manual SAS workflows and reducing reporting cycle time by 60–70%. Established Git DevOps standards and reusable templates across a five-analyst team. Mentored junior analysts through code reviews and knowledge-sharing sessions.
- **Privacy-constrained analytics:** Directed three years of analysis on Washington APCD healthcare claims and DOC release cohorts. Investigated opioid prescription patterns, evaluated PTSD intervention outcomes, and identified key population health metrics despite restricted individual access. Applied reporting thresholds and aggregation to deliver actionable, compliant insights.
- **Experiment design and evaluation:** Designed A/B and quasi-experimental frameworks for DOC programs. Used advanced statistical methods, including chi-square tests and Cramér's V, to distinguish data gaps from quality issues. Prevented misattribution of approximately 45,000 records and ensured targeted remediation.
- Manage five active Agile value domains (63 features), delivering over 20 Power BI dashboards with daily to annual cadences for executive and legislative stakeholders. These dashboards cover operational surveillance, recidivism, misconduct, and community reentry KPIs. Led cross-functional stakeholder interviews to gather requirements, translated business needs into technical features, and facilitated training sessions with non-technical audiences to drive adoption and ensure solutions met operational goals.
- Engineered end-to-end Python pipeline processing 123,291 administrative records across 39 counties and ten-table schemas. Developed composite county cooperation algorithm using weighted metrics and applied chi-square testing to distinguish true data gaps (0.8%) from apparent missingness (45%). Automated identification of 4,107 high-complexity cases and delivered interactive Quarto reports with GeoPandas mapping and NetworkX network analysis.

## **Institute for Learning & Brain Sciences, University of Washington**

2011 – 2023

---

### **Research Science Engineer – Seattle, WA**

- **Predictive ML from high-dimensional data:** Engineered Python pipelines to transform 306-channel MEG time-series into predictive models, including an infant vocabulary predictor ( $r = 0.73$ ,  $p = 0.001$ ) and a reading-skill biomarker ( $r = 0.67$ ,  $p = 1 \times 10^{-6}$ ), both published in leading journals. Applied linear mixed effects, cluster permutation testing, and dimensionality reduction techniques across diverse cohorts.
- **Open-source Python infrastructure:** Co-developed MNE-BIDS (*Journal of Open Source Software*), a Python package standardizing neuroimaging data at scale and adopted internationally by over 15 contributors. Reduced preprocessing setup time per study from days to hours and advanced automated pipelines across Windows, macOS, and Linux.
- **Data pipeline engineering under \$2M+ NIH/NSF funding:** Led development of scalable Python and MATLAB pipelines for a 12-year pediatric cohort program, enabling advanced signal space separation, artifact rejection, and multi-table integration. Produced four peer-reviewed publications using reproducible frameworks.

## **Children's Hospital of Philadelphia, Department of Radiology**      2008 – 2011

---

### **Post-Doctoral Researcher, Lurie Family Foundations MEG Imaging Center – Philadelphia, PA**

- Validated a language impairment classifier (AUC 0.86, sensitivity 82.4%, specificity 71.2%, Cohen's  $d = 3.11$ ) using blinded ROC analysis on 78 pediatric participants. Applied regulatory-grade methods, pre-specified criteria, Bonferroni correction, and linear mixed modeling across 301 observations.
- Designed a four-condition auditory oddball paradigm across two independent cohorts ( $N = 96+$ ), resolving longstanding contradictions by increasing sample sizes five- to eightfold. Applied repeated-measures ANOVA, Morlet wavelet time-frequency analysis, and beamforming to isolate neural signatures.

## **Medical Research Council, Cognition and Brain Sciences Unit**      2007 – 2008

---

### **Visiting Scientist / Post-Doctoral Scientist – Cambridge, United Kingdom**

- Conducted research on speech sound perception and spoken word recognition.
- Collaborated with Prof. Friedemann Pulvermuller on lexical/semantic processing studies.
- Developed paradigms for studying neural plasticity and functional-anatomical neuroimaging.

## **Institute for Biomagnetism and Biosignalanalysis, University Hospital Münster**

2004 – 2007

---

### **Doctoral Fellow / Ph.D. Candidate – Münster, Germany**

- Conducted doctoral research on phonological processing in human auditory cortex using magnetoencephalography.
- Master's thesis advisor for Ludger Elling (2006): "Compensation for regressive place assimilation by the Listener: An MEG Study."
- Published multiple peer-reviewed papers on speech feature encoding using MEG.

## **PROJECTS**

---

### **WADOC NLP Incident Classifier** Production

Deployed a multi-label text classification pipeline processing 110,928 incident records across 73 categories in production. Built scikit-learn pipeline with TF-IDF features (15,000 vocabulary, unigrams and bigrams) and a MultiOutputClassifier, achieving 75.6% micro-F1 and 88.3% precision – a 30-fold improvement over baseline. Integrated human-in-the-loop review system for continuous model refinement and maintained deployment stability.

Python · scikit-learn · pandas · TF-IDF · Oracle/SQL · NLP · production ML

### **WA DOC Out-of-Custody Analytics Pipeline** Production

Engineered end-to-end Python pipeline processing 123,291 administrative records across 39 counties and ten-table schemas. Developed composite county cooperation algorithm using weighted metrics and applied chi-square testing to distinguish true data gaps (0.8%) from apparent missingness (45%). Automated identification of 4,107 high-complexity cases and delivered interactive Quarto reports with GeoPandas mapping and NetworkX network analysis.

Python · pandas · GeoPandas · NetworkX · SciPy · T-SQL · Quarto

## MNE-BIDS – Open-Source Python Package Open Source

Co-developed MNE-BIDS, a Python package published in the *Journal of Open Source Software* (DOI 10.21105/joss.01896), standardizing electrophysiological data and enabling reproducible pipelines. Collaborated with over 15 international contributors and secured NIH, NIMH, and ERC funding. Reduced data preparation time from hours to minutes and ensured robust testing across Windows, macOS, and Linux.

Python · BIDS · MNE-Python · CI/CD · GitHub · Cross-platform testing

## Automaticity in the Reading Circuitry Research

Developed Python-based ML pipeline for analyzing speech-related neural responses in children (N=42, ages 7–12). Applied machine learning algorithms including neural networks and dimensionality reduction (PCA). 3D statistical modeling of dense-array timeseries data published in *Brain and Language*.

Python · MNE-Python · PCA · MEG · scikit-learn

## Speech Discrimination Biomarkers for Language Impairment Research

Developed ML classification system using nonparametric linear mixed modeling to analyze speech discrimination patterns in children with autism spectrum disorders (N=51, ages 6–15). Achieved clinically-relevant diagnostic accuracy (AUC 0.86) for language impairment detection. Published in *Biological Psychiatry*.

Python · MATLAB · ROC analysis · linear mixed models · MEG

## PUBLICATIONS

---

1. Bosseler, A.N., Clarke, M., **Tavabi, K.**, Larson, E.D., Hippe, D.S., Taulu, S., Kuhl, P.K. (2021). Using Magnetoencephalography to Examine Word Recognition, Lateralization, and Future Language Skills in 14-Month-Old Infants. *Developmental Cognitive Neuroscience*, 47, 100901.
2. Joo, S.J., **Tavabi, K.**, Caffarra, S., Yeatman, J.D. (2021). Automaticity in the Reading Circuitry. *Brain and Language*, 214, 104906.

3. Clarke, M., Larson, E., **Tavabi, K.**, Taulu, S. (2020). Effectively Combining Temporal Projection Noise Suppression Methods in Magnetoencephalography. *Journal of Neuroscience Methods*, 341, 108700.
4. Appelhoff, S., Sanderson, M., Brooks, T., van Vliet, M., Quentin, R., Holdgraf, C., Chaumon, M., Mikulan, E., **Tavabi, K.**, et al. (2019). MNE-BIDS: Organizing Electrophysiological Data into the BIDS Format. *Journal of Open Source Software*, 4(44), 1896.
5. Roberts, T.P.L., Cannon, K.M., **Tavabi, K.**, Blaskey, L., Khan, S.Y., Monroe, J.F., et al. (2011). Auditory Magnetic Mismatch Field Latency: A Biomarker for Language Impairment in Autism. *Biological Psychiatry*, 70(3), 263–269.
6. **Tavabi, K.**, Embick, D., Roberts, T.P.L. (2011). Spectral–Temporal Analysis of Cortical Oscillations during Lexical Processing. *NeuroReport*, 22(10), 474–478.
7. **Tavabi, K.**, Embick, D., Roberts, T.P.L. (2011). Word Repetition Priming–Induced Oscillations in Auditory Cortex. *NeuroReport*, 22(17), 887–891.
8. **Tavabi, K.**, Elling, L., Dobel, C., Pantev, C., Zwitterlood, P. (2009). Effects of Place of Articulation Changes on Auditory Neural Activity. *PLoS ONE*, 4(2), e4452.
9. **Tavabi, K.**, Obleser, J., Dobel, C., Pantev, C. (2007). Auditory Evoked Fields Differentially Encode Speech Features. *European Journal of Neuroscience*, 25(10), 3155–3162.
10. Villablanca, J.R., Schmanke, T.D., Crutcher, H.A., Sung, A.C., **Tavabi, K.** (2000). The Growth of the Feline Brain from Late Fetal into Adult Life. *Developmental Brain Research*, 122(1), 21–33.

## EDUCATION

---

**PhD in Cognitive Psychology / Neuroscience** – University of Münster, Germany  
2007

---

**MSc in Psychology** – University of Oregon, USA

2004

---

**BSc in Physiological Sciences** – University of California, Los Angeles, USA  
2001

---

## CERTIFICATIONS

---

- **2026:** IBM Deep Learning Professional Certificate (Coursera)
- **2026:** Machine Learning with Python – IBM AI Engineering Professional Certificate (Coursera)
- **2024:** Supervised Machine Learning: Regression and Classification (Coursera)
- **2023:** Bayesian Statistics: From Concept to Data Analysis (Coursera)
- **2023:** Certified SQL Developer (W3Schools)
- **2022:** Statistical Learning (edX)
- **2012:** Elekta Neuromag® MEG Advanced Program (Helsinki, Finland)

## SKILLS

---

**ML & Statistics:** Python, scikit-learn, pandas, NumPy, SciPy, PyTorch, TensorFlow, logistic regression, SVM, random forest, linear mixed effects, ROC/AUC, causal inference, A/B testing, chi-square, ANOVA, Bayesian statistics, effect sizes

**Speech & Language Processing:** Digital signal processing, audio signal analysis, NLP, TF-IDF vectorization, text classification, time-series analysis, spectral-temporal analysis, multi-sensor data fusion

**Healthcare Data:** APCD (All-Payer Claims Database), insurance claims analysis, population health cohort design, clinical biomarker validation, IRB protocols, privacy-constrained analytics

**Data Engineering:** SQL (T-SQL, Oracle), ETL pipeline design, Power BI (semantic models, DAX), Power Automate, Azure DevOps, Git, Quarto, R Markdown

**Languages:** Python, R (ggplot2, lme4, Tidyverse), MATLAB, SQL, Shell (Bash/Zsh), LaTeX

**Platforms:** Linux, macOS, Windows, Azure DevOps, Git, GitHub, VS Code

**Domain:** Neuroscience, pediatric clinical research, correctional health, Medicare Advantage / managed care analytics, geospatial analytics

## GRANTS & AWARDS

---

- **2010–2012:** National Institute of Health Loan Repayment Program – NIDCD funding for extramural clinical research
- **2007–2008:** Post-Doctoral Fellowship "NESTCOM: What it means to communicate"  
– NEST-2005-PATH HUM
- **2004–2007:** Doctoral Fellowship, Faculty of Medicine, University Hospital Münster, Germany

[Download PDF version](#)