MAY MERKLE-TAN, PHD

DATA SCIENTIST

Trained as a neuroimaging scientist studying brain functions in health and disease, May thrives in the interdisciplinary process of discovery and translating findings into data-stories. She is committed to integrating best practices in machine learning, statistical methods, and human interaction to derive actionable insights. May grew up in Singapore, is educated in the UK and Germany, and is a US permanent resident (green-card holder).

hengrumay@gmail.com
hengrumay.github.io
917-932-6639
hengrumay
hengrumay

COMMUNICATION | IMPACT

LINK TO PUBLICATIONS (10) & TALKS (27): HTTP://ORCID.ORG/0000-0003-2109-0781

Involved in knowledge sharing (developing enterprise analytics repository; communicating research; giving lectures on neuroimaging methods; scientific peer-reviewing; initiated and ran journal club for researchers), mentoring and outreach (supervising STEM students); participating in public data dives (e.g. DataKind | CapitalOne DevExchange | BBC 100 Women 2017 Challenge)

EDUCATION

UNIVERSITY OF EDINBURGH, UK
PhD Cognitive Neuroimaging
Research supported by The Carnegie Trust and conducted in collaboration with the Brain Sciences Center, University of Minnesota, Minneapolis VA Medical Center, USA

INTERNATIONAL MAX PLANCK RESEARCH SCHOOL, UNIVERSITY OF TÜBINGEN, GERMANY
MSc Neural & Behavioural Sciences
Thesis: Investigated somatosensory discrimination processes using evoked magnetic field analysis and equivalent dipole source estimation of neural signals

UNIVERSITY OF EDINBURGH, UK
MA (Hons) Psychology
Honors Concentration: Cognitive Neuroscience, Neuropsychology | Additional Courses: Engineering Mathematics, Linguistics

SKILLS

PROGRAMMING	Python, R, Matlab
STATISTICAL	(Non- &) Parametric Statistics, Test-Retest Reliability, Matched-Pairs, SPSS
MACHINE LEARNING | APPLICATIONS	Regression, Time-Frequency, Time-Series, Classification & Clustering Approaches, Natural Language Processing (NLP), Neural Networks
WEB SERVICES | DATABASE	AWS, SQL
WEB-BASED EXPERIENCE	Git & Github, HTML, CSS, Flask, Beautifulsoup
GRAPHICS & VISUALIZATION	Adobe Illustrator, GIMP, seaborn | ggplot2, D3.js
RESEARCH	Experimental Design, User-Interface, Data Acquisition, Study Management, Publication & Communication, Scientific Peer Reviewing

EXPERIENCE

Data Scientist | Creative
INDEPENDENT CONTRIBUTOR | FOSSICKLABS LLC, DENVER, CO — Oct. 2017 - Current

- Consult on data science projects; Author technical/white-paper(s) on the application of Machine Learning/NLP approaches in industry, e.g. Healthcare; Beta-testing data science curriculum|assessment UXI; Develop prospective customer experience trouble-shooting analytics and applications for Fortune50 company
- Collaborate with industry partners to develop and build out data science applications

Data Science | Technical Consultant
COMCAST | EBI--DATA PRODUCTS AND ANALYTICS, ENGLEWOOD, CO — March 2018 - June 2018

- Developed data science process, analytical tools in at least half the time typically anticipated: Integrated data from multiple databases (e.g. Teradata); Built machine learning pipelines for assessing and tuning multiple classification models; Extrapolated insights from big data modeling that links machine data with customer feedback to aid strategic decision-making in improving customer experience and cost-management

Data Scientist | Independent Contributor
METIS, NYC, USA — Oct. 2016 - July 2017

Integrating research, machine-learning, and statistical tools to derive data-driven insights from full-stack data science projects such as:

- Integrated weather and air quality data to predict (83% accuracy) the frequency of asthma patients' visit to the emergency department using gradient boosted trees; rendered a prototype D3.js calendar application
- Applied Natural Language Processing to develop a "Recipe-Difficulty-Tagger" using structured learning and topic modeling (e.g. LDA), engineered features for classifying recipe preparation difficulty (~86% precision & recall) and created a web-app to offer recipe alternatives based on ingredient topic similarity and categorized by difficulty
- Developed and trained convolutional neural networks (CNN) using Keras with TensorFlow to detect buildings in a highly unbalanced set of proprietary satellite images and generated interactive visualizations of the test image classification outcomes (~93% accuracy) using geospatial information embedded in the GeoTIFFs; Further explored image segmentation using computer vision techniques to delineate buildings in images

Visiting & External Research Affiliate
CENTRE FOR COGNITIVE NEUROIMAGING, UNIVERSITY OF GLASGOW, UK — Oct. 2015 - March 2017

- Continued collaboration on cross-frequency analysis of brain signals to uncover how dynamic sensory information could be neurally propagated to drive action intention

Research Fellow
INSTITUTE OF BEHAVIOURAL NEUROSCIENCE, UNIVERSITY COLLEGE LONDON, UK — Feb. 2012 - Aug. 2012

- Developed a virtual reality navigation paradigm (using object-oriented Python) for studying (both behaviorally and neurally) how navigators acquire their understanding of an environment and respond to unanticipated changes

Senior Research Associate
CENTRE FOR COGNITIVE NEUROIMAGING, UNIVERSITY OF GLASGOW, UK — Sept. 2009 - Sept. 2015

- Led the examining of test-retest reliability and differential neural oscillatory signals in health and neurological/psychiatric disease; assessed the suitability of neural oscillatory parameters for potential biomarkers in early detection, prognosis, and/or remediation; Demonstrated high repeatability of signals derived from sensors and approximated source estimations facilitating practical and analytical recommendations for large-scale applications
- Involved in projects funded by the Economic and Social Research and the Medical Research Councils, UK

Research Scientist
BRAIN SCIENCES CENTER, UNIVERSITY OF MINNESOTA, USA — March 2008 - Sept. 2009

- Assessed the synchronous neural activity in healthy controls and veterans with Post-traumatic Stress Disorder using magnetoencephalography (MEG) for potential diagnostic and therapeutic applications; Contributed to the development of open-loop classification of patient vs. healthy controls -- mimicking the scenario where you do not necessarily have a 'fixed' training and test dataset for diagnosis; yielded high sensitivity & specificity (~90%)