

# Tanya Wen, PhD

---

Email: [2tanyawen@gmail.com](mailto:2tanyawen@gmail.com) | +1-919-450-5465 | Website: [tanya-wen.github.io/](https://tanya-wen.github.io/) | [linkedin.com/in/tanya-wen-phd/](https://linkedin.com/in/tanya-wen-phd/)

## Professional Summary

Industry scientist with proven skills of analytics, experimentation, and infrastructure optimization at large-scale social media platform serving billions of users, translating insights into product and technical decisions. Drove data-informed decisions across video, image, and audio to improve system reliability, reduce costs, and enhance user experience. Cross-functional collaborator with expertise in both Data Science and UX Research.

## Work Experience

### Data Scientist | Meta, Family of Apps

Aug 2025 – Present

- Conducted a 4-part series benchmarking analysis of internal and external messaging apps using production data and lab experiments to understand topline engagement patterns, send-to-sent reliability, render quality, and user experience across 6 media types to provide actionable product improvements.
- Identified opportunity of using additional features to personalize the current ML model of media uploads (optimizing reliability, latency, and visual quality tradeoffs). Ran simulations of different scale factors in the updated model to size the impact on upload bytes across user cohorts and inform experimentation.
- Conducted comprehensive analysis of video prefetching across Facebook and Instagram; reviewed 50+ historical experiments to identify 7 key optimization levers; ran 10+ A/B testing experiments to fine-tune the prefetch infrastructure to improve video buffering and egress usage.
- Root caused 4 major regressions in media reliability, 6 common error codes, enabling engineering to ship targeted fixes. Identified 3 logging bugs that underestimated feature improvement opportunities, partnered with engineers to land code fixes and reprioritize overlooked areas on the team's roadmap.
- Ownership of forecasting analyses to predict the surge of media sends on Messenger in preparation for regional and global holiday surges, supporting XFN, server, and oncall teams in operational readiness.
- Built an automatic metric collection scheduler that matches key metrics to experiment domains, alongside 5 dataswarm pipelines which powered the Media Foundation team's dashboard for tracking launched and active experiments with regression monitoring.
- Built 8 analytics agent recipes with context of core tables, SQL query patterns, decision logic, response formatting, and validation checks. Accelerated regression investigation from days to ~30 mins, automated weekly metrics reporting, and facilitated launch reviews.
- Supported engineering teams by promoting best practices in data science and held office hours, covering metric selection, results interpretation, and review of 60+ A/B tests, to ensure rigorous and reliable decision-making. Identified Media Messaging PREQ improvements drove +30M daily sends and +5M daily sessions; set PREQ goals during roadmap planning and designed holdout structure with XFN teams.

### Applied Perception Scientist | Meta Reality Labs (contract via Magnit)

Mar 2024 – Aug 2025

- Led the experimental design, analysis, and results documentation of four KPI metrics to A/B test the effect of audio personalization on the Quest 3 VR headset across 5 studies, which helped inform decision-making on our team's technology roadmap. This also resulted in two publications and open-source contributions.
- Led our team's UXR study to evaluate user value of a newly developed audio renderer when conversing with others in mixed reality in an avatar-based context. Used a mixture of quantitative methods including Bradley-Terry model for ranking and qualitative thematic analysis of user feedback to inform feature value.
- Co-led a series of end-to-end onsite user studies to evaluate speech intelligibility on Meta Ray-Ban glasses, measuring psychophysics thresholds and qualitative ratings, which contributed to the Conversation Focus feature, announced at Connect 2025.

- Initiated a new workstream of online data collection, developed a web-based tooling infrastructure using Prolific, Qualtrics, and Amazon S3 to enable rapid collection of large-scale user data, which increased the team's efficiency by at least 30x and reduced the cost of data collection by ~80%. This was demoed to stakeholders across XFN teams and became a main data collection method in our team.
- Collaborated with cross-functional teams, including supervising research assistants in initiating data collection, working with SWEs to debug Unity Apps, and discussing study results with research scientists.

**Research Scientist | Naval Health Research Center (contract via Leidos) Oct 2022 – Mar 2024**

- Led analysis of EEG recordings in 50+ traumatic brain injury patients and healthy controls, utilizing ERP analysis, k-means clustering, and exploratory data analysis to identify biomarkers. Results were included in our annual report and conference.
- Designed and programmed four different virtual reality tasks in Unity / C#, interact with hardware (including light sensor, sound sensor, IMU), and timestamp events to send to Tobii eye-tracking glasses to sync across multiple devices.
- Led the data analysis and study documentation of an interactive virtual reality experiment to measure gaze stability, fixations, and saccades using a wearable eye-tracker and infrared camera. Developed custom analysis scripts using computer vision algorithms in OpenCV and pre-trained neural networks.

**Postdoctoral Associate | Center for Cognitive Neuroscience, Duke University Oct 2019 – Oct 2022**

- Demonstrated cognitive training and transfer learning in a series of three experiments using hierarchical Bayesian reinforcement learning models.
- Designed and published three innovative studies that generated novel discoveries in human cognition (transfer learning, temporal memory, and relative effort).
- Designed and programmed 20+ web-based experiments, and collected data from over 800+ participants on Amazon Mechanical Turk.
- Developed fMRI processing pipeline for the lab utilizing the Duke Compute Cluster with SLURM and Python scripts for conducting generalized linear models, increasing the reliability and fidelity of the analysis with automation, and making the learning process more accessible for students.

## Skills

**Quantitative Skills:** inferential statistics (A/B testing, t-test, ANOVA); regression; linear mixed effects modeling; exploratory data analysis; supervised learning (logistic regression, random forest, GBDT, SVM); causal inference (DiD, propensity score matching, instrumental variables); forecasting; opportunity sizing.

**User Research:** experimental design; research planning; metrics design; virtual reality, mixed reality; eye-tracking; psychophysics; qualitative surveys; user interviews; online-studies; in-person studies; participant recruitment; IRB application; prototyping; hardware UX; thematic analysis; sentiment analysis.

**Programming & Engineering:** Python (numpy, pandas, SciPy, Scikit-learn, Matplotlib, seaborn); SQL (Hive, Presto); MATLAB; JavaScript; Unity, C#; R (lme4, tidyverse, ggplot2); Git and Github; data pipelines; Claude.

## Education

**PhD in Medical Science**

MRC Cognition and Brain Sciences Unit, University of Cambridge, UK

**Oct 2015 – Sept 2019**

**Bachelor of Science, Double Major**

Department of Psychology, National Cheng Kung University, Taiwan  
 Department of Life Sciences, National Cheng Kung University, Taiwan

**Sept 2011 – May 2015**

## Publications

See my Google scholar page: <https://scholar.google.com/citations?user=pegh2ooAAAAJ>