Yi Chern Tan

yichern.tan[at]aya.yale.edu | linkedin.com/in/yicherntan | github.com/yichern | google scholar | website: yich.io

Experience

Cohere

Post-Training Co-Lead, Generative Modeling

- Co-led the post-training program across ~60 research scientists and engineers, led the implementation of a new agent and reasoning prompt template, alignment of agentic post-training with other domains, and the collaboration for quantization and serving, which culminated in the launch of Command A (1313 Elo, 13th at launch, DeepSeek v3 performance), serve-able with 2 H100s, and earlydelivery with key commercial partners, as well as Command R7B (top of OpenLLM leaderboard at weight class)
- Scientific advisor for the merging of post-trained checkpoints across domains at SFT and preference tuning, led to merging as the key innovation of a scalable post-training strategy across domains as documented in the Command A technical report

Senior Member of Technical Staff, Generative Modeling

- Core post-training contributor for Command R and Command R+, led the implementation of prompt template for instruction and preference data, packing and masking of chat data, team processes for cyclical model training, as well as ran thousands of experiments and ablations for the entire post-training pipeline, which culminated in the launch of Command R (1146 Elo, 10th at launch) and Command R+ (1194, 6th at launch, just below Claude 3 Sonnet)
- Conceptualised Cohere's internal annotation platform needs and tech lead for a team of 8 across engineering, design and product to build the platform from 0 to 1, reducing spend on external annotation vendors drastically and improving post-training human annotation quality via in-house controls, data review, and bespoke annotation pipelines

Waymo LLC

Machine Learning Engineer, Simulation and Safety Evaluation

- Improved automated data refresh, training and evaluation infrastructure for a classifier that scores the self-driving car's likelihood of a realistic collision with other road users in simulation, decreasing experiment iteration time from ~ 5 days to ~ 2 days
- Implemented feature engineering and architectural changes (e.g. mixture of experts) to the collision likelihood model, improving P@R95 by ~4% overall and ~22% for vulnerable road users, reducing human-in-the-loop triage load by ~10%.
- Proposed and launched a new model to classify safety-relevant incidents (e.g. tailgating, lane changes), improved accuracy by ~20%.

Facebook, Inc.

Software Engineering Intern, Instagram

- Developed high-performant text classifiers for detecting violating comments (toxicity, bullying) with a new contextual encoder, distributed learning, and knowledge distillation; increased calibrated recall by ~70% at ~40% lower latency across 9 languages
- Implemented MLOps tooling for model training at scale (automated training pipelines, ~billions of comments), evaluation (score visualization) and deployment (publishing workflows) of classifiers; increased model iteration speed and robustness
- Secured the highest tier full-time return offer (top ~5%, "greatly exceeds expectations" performance review)

TAIGER

Applied Research Scientist Intern, NLP

- Built pipeline for identity card image segmentation and extraction, increasing AP by >70% (TensorFlow, OpenCV)
- Prototyped a paraphrase generation technique for augmenting an English creole dataset of municipal queries using unsupervised machine translation, increasing BLEU by >2 (PyTorch)

Representative Research and Peer-Reviewed Publications

- '23 '25 (Lead): Command A, Command R, Command R+; (Contributor): Aya 23, Aya Expanse
- NeurIPS '19 (1st author, spotlight) Assessing Social and Intersectional Biases in Contextualized Word Representations
- ACL '19 (co 1st author, BlackBoxNLP workshop) Open Sesame: Getting Inside BERT's Linguistic Knowledge
- ACL '19 (4th author) SParC: Cross-Domain Semantic Parsing in Context
- ACL '20 (3rd author) ESPRIT: Explaining Solutions to Physical Reasoning Tasks
- ICLR '21 (4th author) GraPPa: Grammar-Augmented Pre-Training for Table Semantic Parsing
- Affiliations and Collaborators: Max Bartolo, Patrick Lewis, Phil Blunsom, Ed Grefenstette, Aurélien Rodriguez, Sara Hooker (Cohere), Elisa Celis (Controlling Bias in AI Group), Robert Frank (Computational Linguistics at Yale Lab), Tao Yu, Rui Zhang, Dragomir Radev (Language Information and Learning at Yale Lab), Nazneen Rajani and Xi Victoria Lin (then Salesforce Research)
- Academic Reviewing: EMNLP '21 '22, NeurIPS '21 (Outstanding Reviewer Award) '25, ICLR '22 '25, ICML '22 '24

Education

Yale University

Bachelor of Science (CGPA: 3.98), Summa Cum Laude Double Major in Computer Science (GPA: 4.00) and Ethics, Politics, Economics (GPA: 3.97)

New Haven, CT Aug '16 - May '20

London, UK; Singapore June '24 - present

Sep '22 - June '24

Menlo Park, CA

Mountain View, CA

Mar '22 - Sep '22

May '19 - Aug '19

Singapore

Dec '18 - Jan '19

• Honors: Phi Beta Kappa (top 5% of cohort), Benjamin Franklin College Valedictorian

• Selected Coursework: NLP (PhD level), ML, Data Science Ethics, Normative Ethics, Classical AI, Robotics, Operating Systems