Tai Dang

tai-dang11.github.io / taitdang11@gmail.com

Education & Training

Stanford University *Visiting Researcher*
Advisors: Thang Luong (Google Deepmind) & Jeff Glenn (Stanford Medicine)

University of Massachusetts Amherst *B.S. in Computer Science, May '24*
Advisors: Bruno Castro da Silva (Computer Science)

Research

My research focuses on generative modeling and decision-making, particularly diffusion models and reinforcement learning, with applications in computational biology. It supports discovery by integrating generative and decision-making methods. Below are selected papers:

1. **Dang** et al. High-Fidelity Molecular Structure Prediction via Reinforcement Learning. *Preprint'26*
   - Achieved SOTA in structure prediction fidelity and affinity via RL with physics-based rewards.
2. **Dang** et al. Drug Discovery with Expert Preferences. *Preprint '25*
   - Recovered 16/37 EGFR and 37/58 DRD2 drugs from 100K ligands via chemist-guided screening.
   - *Paper:* arXiv, *Code:* tai-dang11/cheapvs
3. **Dang** et al. Enriching Biomedical Knowledge for Low-resource Language Through Translation. *EACL*
   - SOTA in Vietnamese biomedical benchmark and high-quality Vietnamese MedNLI dataset.
   - *Paper:* eacl, *Code:* vietai/ViPubmed
4. MTet: Multi-domain Translation for English and Vietnamese. *Preprint*
   - *Details:* SOTA in English-Vietnamese translation and high-quality multi-domain bilingual corpus.
   - *Paper:* arXiv, *Code:* vietai/mTet
5. AURORA-M: Open Source Continual Pre-training for Multilingual Language and Code *coling'25*
   - Developed a 15B open-source multilingual model continually pre-trained on code and text.
   - *Paper:* coling, *Model:* huggingface.co/aurora-m
6. Gathering Context that Supports Decisions via Entropy Search with Language Models. *Preprint'26*
   - Closed 85% of performance gap to fully-informed agents via uncertainty-driven information seeking.

Experience

1. Stanford University – *Visiting Researcher '24–Present*
   - Post-trained AlphaFold 3 via reinforcement learning, SOTA on structure fidelity.
   - Achieved optimal drug screening on large libraries via Bayesian optimization.
2. Umass Amherst – *Research Assistant '23*
   - Engineered a multi-modal retrieval system for Outside-Knowledge Visual QA.
   - Revealed higher discount factors increase policy specialization in the OLS Convex Coverage Set.
3. Ontocord – *Research Intern '23*
   - Distilled 7B LLM model to **5x** smaller size while maintaining performance parity.
   - Developed open-source Vietnamese LLM by processing 1TB of data.
4. EOG Resources – *Software Engineer Intern '23*
   - Built graph-based visualization software to streamline complex data analysis.
   - Migrated repositories to GitHub Actions and implemented OIDC authentication.
5. VietAI – *Research Intern '22*
   - Developed SOTA En-Vi model and improved Biomedical NMT via self-training.
   - Drove 6% BLEU improvement in Biomedical NMT via self-training and released Vi-MedNLI dataset.
6. FPT Software – *Research Intern '21*
   - Developed interactive frontend for internal AI model services.

Project

1. Multi-Objective GFlowNet for Drug Design
   - Generated diverse, synthesizable molecules with optimized affinity and ADMET via SynFlowNet.

Honors & Funding

Paper Awards: ICLR'25 Workshop Spotlight, ICML'25 Workshop Spotlight
Grants: Google-HAI Grant '24 ($90,000), Google-HAI Grant '25 ($100,000), Stanford Marlowe Grant.

Skills

Frameworks: PyTorch, Flax, JAX, TensorFlow, Hugging Face, Flask, Node.js, Neo4j
Languages: Python, Java, JavaScript, C/C++, SQL
Tools: Git, Linux, GCP, Slurm, Docker, GitHub, Kubernetes