葉適穎 Shih-Ying Yeh

a.k.a. Kohaku BlueLeaf, Hsinchu, Taiwan

Personal Profile

- **Contact:** kohaku@kblueleaf.net
- **Personal Homepage:** https://kblueleaf.net
- **GitHub Profile:** https://github.com/KohakuBlueleaf
- **HuggingFace Profile:** https://huggingface.co/KBlueLeaf
- **Google Scholar:** https://scholar.google.com/citations?user=XmW5oGIAAAAJ
- **LinkedIn:** https://www.linkedin.com/in/kblueleaf

Education & Career History

Master student Computer Science, National Tsing Hua University 2025 – Present

- Advisor: Shang-Hong Lai prof.

Undergrad student Computer Science, National Tsing Hua University 2021 – 2025

ML/NN Engineer Comfy Org, core developer of ComfyUI 2025/03 – Present

- Created and maintaining the whole weight adapter system (LoRA/LoHa/LoKr/DoRA/OFT/BOFT), lora trainer nodes, dataset nodes and FP4/FP8/FP16 training.

Founder Kohaku Lab 2025/04 – Present

- Virtual open source research lab. KohakuTerrarium, KohakuTPU, KohakuHub, KohakuRiver, KohakuRAG and more.

ML Engineer PT Apple Inc. 2024/07 – 2024/11

- Graph-Based Captioning dataset. Published on arXiv and Apple official github.

Consultant Element, Inc. 2023/09 – 2024/07

- Personalized Text-to-Image experience and video-to-video features.

Full Stack Web Developer Fungs Design 2021/08 – 2022/02

Expertise

- Generative Models: Text-to-Image Models, LLM
- Parameter-Efficient Fine-Tuning
- Agent System
- Compiler, IR and programming language design
- AI accelerator design (KohakuAccel framework), logic circuit design, FPGA
- ML Infrastructure: cluster manager, model hub, experiment tracking
- Deep Learning, Computer Vision
- Web development

Publications

Google Scholar: 229 citations, h-index 4 (2026/10)

1. **TT-VidT: Decoupling the Temporal Axis for Efficient Motion-Centric Video Pretraining**
   Shih-Ying Yeh, Daniel Z. Kaplan, Xuehai Wang, Fu-En Yang, Min-Hung Chen, Shang-Hong Lai (2026)
   NeurIPS 2026 | arXiv:2609.33419
   - Accepted by NeurIPS 2026 main track.

2. **Trinity: One Differentiable Physics for Training, Refining and Scoring Generative Floorplanners**
   Shih-Ying Yeh, Tzu-Sian Wang, Xuehai Wang, Jia-Hua Lee, Daniel Z. Kaplan, Ming-Qi Xu, Wuqian Tang, Chun-Yao Wang, Shang-Hong Lai, Chun-Yi Lee (2026)
   arXiv | arXiv:2610.04957
   - Equal contribution with Tzu-Sian Wang. Code and models open sourced.

3. **World Editing: Intervening on Executable Worlds at Increasing Depth**
   Max Ku, Nok-Kan Law, Yu-Chien Tang, **Shih-Ying Yeh**, ..., Ho Kei Cheng (18 authors) (2026)
   arXiv | arXiv:2610.02331

4. **ReactVAU: A Slow-Fast Decoupled Framework for Streaming Video Anomaly Understanding**
   Chia-Hui Chen, **Shih-Ying Yeh**, Fu-En Yang, Min-Hung Chen, Shang-Hong Lai (2026)
   ECCV 2026 | arXiv:2609.07941
   - Accepted by ECCV 2026.

5. **TIPO: Text to Image with Text Presampling for Prompt Optimization**
   Shih-Ying Yeh, Yi Li, Sang-Hyun Park, Giyeong Oh, Xuehai Wang, Min Song, Youngjae Yu, Shang-Hong Lai (2026)
   ICLR 2026 | OpenReview | arXiv:2411.08127
   - Originated as my undergraduate project (advisor: Shang-Hong Lai prof.). Later collaborated with Yonsei University's laboratories and OnomaAI company.
   - Related open-source models have achieved over 3M downloads, related open source projects have received over 700 stars on GitHub in total.

6. **ImagenWorld: Stress-Testing Image Generation Models with Explainable Human Evaluation on Open-ended Real-World Tasks**
   Samin Mahdizadeh Sani, Max Ku, Nima Jamali, ..., **Shih-Ying Yeh**, ..., Wenhu Chen (26 authors) (2026)
   ICLR 2026 | OpenReview | arXiv:2603.27862

7. **KohakuRAG: A Simple RAG Framework with Hierarchical Document Indexing**
   Shih-Ying Yeh, Yueh-Feng Ku, Ko-Wei Huang, Buu-Khang Tu (2026)
   arXiv | arXiv:2603.07612
   - #1 in Kaggle WattBot 2025 competition.

8. **Home-made Diffusion Model from Scratch to Hatch**
   Shih-Ying Yeh (2025)
   arXiv | arXiv:2509.06068
   - Tech report of HDM project. T2I model trained from scratch with 4 RTX5090 at home.

9. **Graph-Based Captioning: Enhancing Visual Descriptions by Interconnecting Region Captions**
   Yu-Guan Hsieh, Cheng-Yu Hsieh, **Shih-Ying Yeh**, Louis Béthune, Hadi Pouransari, Pavan Kumar Anasosalu Vasu, Chun-Liang Li, Ranjay Krishna, Oncel Tuzel, Marco Cuturi (2024)
   arXiv | arXiv:2407.06723
   - Collaborated with Apple Inc.

10. **Navigating Text-To-Image Customization: From LyCORIS Fine-Tuning to Model Evaluation**
    Shih-Ying Yeh, Yu-Guan Hsieh, Zhidong Gao, Bernard B W Yang, Giyeong Oh, Yanmin Gong (2024)
    ICLR 2024 | OpenReview | arXiv:2309.14859
    - ICLR 2024 Poster, 203 citations.
    - Related open-source projects got over 3900 stars on GitHub in total.

Open-Source Projects / Contributions / Models

Total 8k+ stars on GitHub, 6.8M+ PyPI downloads, 3.8M+ HuggingFace downloads (2026/10)

1. **LyCORIS** - Lora beYond Conventional methods, Other Rank adaptation Implementations for Stable diffusion.
   - GitHub Link: https://github.com/KohakuBlueleaf/LyCORIS
   - Open-source implementation of paper "Navigating Text-To-Image Customization: From LyCORIS Fine-Tuning to Model Evaluation."
   - Besides the algorithms introduced by LyCORIS, this project also implemented other frontier PEFT method, such as: GLoRA, DoRA, Diag-OFT, BOFT, $(IA)^3$
   - Has been widely used by Text-to-Image model and LLM community: have been included in Stable-Diffusion-Webui, sd-scripts, SimpleTuner, PEFT
   - Total over 3900 stars on GitHub (LyCORIS, LoCon and sd-webui extensions), over 3.5M downloads on PyPI.
2. **KGen** - A System for Prompt Generation to Improve Text-to-Image Performance.
   - GitHub Link: https://github.com/KohakuBlueleaf/KGen
   - Open-source implementation of paper "TIPO: Text to Image with Text Presampling for Prompt Optimization." and its prototype "DanTagGen".
   - z-tipo-extension: open-source implementation to utilize TIPO in sd-webui or ComfyUI.
   - Total over 700 stars on GitHub, over 220k downloads on PyPI.
3. **HDM** - Home-made Diffusion Model, T2I model pretrained on consumer level hardware.
   - GitHub Link: https://github.com/KohakuBlueleaf/HDM
   - XUT (Cross-U-Transformer) arch with TREAD acceleration and shifted square crop recipe.
   - 340M model trained with 4 RTX5090 at home, cost around 500~600 USD. ComfyUI loader (HDM-ext) provided.
   - Over 200 stars.
4. **KohakuTerrarium** - General purpose AI agent framework.
   - GitHub Link: https://github.com/Kohaku-Lab/KohakuTerrarium
   - Built-in tools, sub-agents, persistent sessions, TUI and web UI. Same agent can be used as library in few lines.
   - kt CLI, official creature pack (kt-biome) and package marketplace (TerrariumMarket).
   - Over 500 stars in 8 months, over 25k downloads on PyPI.
5. **KohakuTPU [WIP]** - AI Accelerator implementation with Xilinx FPGA.
   - GitHub Link: https://github.com/KohakuBlueleaf/KohakuTPU
   - Built on KohakuAccel: open platform for building FPGA accelerators, includes RTL, compiler and driver.
6. **Togeari** - Out-of-order dataflow machine for KIR graph programs.
   - GitHub Link: https://github.com/KohakuBlueleaf/Togeari
   - KIR (KohakuNodeIR) is a Turing complete graph program IR. Togeari runs thousands of KIR jobs at once on pluggable engines, batches calls across jobs.

7. **KohakuGBC** - Togeari based implementation of Graph-Based Captioning (GBC).
   - GitHub Link: https://github.com/KohakuBlueleaf/KohakuGBC
   - Builds graph annotation of each image (entities, regions and their relations) with VLM, OWLv2/SAM 3 and EmbeddingGemma. Implementation of our GBC paper.
8. **KohakuMossTrainer** - Finetuning framework for MOSS-Image.
   - GitHub Link: https://github.com/KohakuBlueleaf/KohakuMossTrainer
   - All LyCORIS algorithms and full finetune at native resolutions, from 6GB cards to multi-GPU nodes. Over 130 stars in 10 days.
9. **PixelOE** - Detail-Oriented Pixelization based on Contrast-Aware Outline Expansion.
   - GitHub Link: https://github.com/KohakuBlueleaf/PixelOE
   - Contrast-Aware Outline Expansion: Use contrast (difference between median and mean value of this patch) as information to blend eroded and dilated images.
   - Near SoTA quality of pixelized images, without any Neural Network. Over 530 stars, over 2.4M downloads on PyPI.
10. **KohakuHub** - An open sourced alternative for HuggingFace, fully compatible with HuggingFace Hub, Transformers and Diffusers.
    - GitHub Link: https://github.com/KohakuBlueleaf/KohakuHub
    - Git-like versioning for models and datasets. Over 250 stars.
11. **KohakuRAG** - #1 in Wattbot2025. Domain-agnostic Retrieval-Augmented Generation (RAG) framework designed for production use.
    - GitHub Link: https://github.com/Kohaku-Lab/KohakuRAG
    - Hierarchical document indexing, context-aware search, single-file SQLite storage (KohakuVault).
12. **KohakuRiver** - Lightweight cluster manager that turns a small fleet of nodes into one powerful computer using Docker.
    - GitHub Link: https://github.com/KohakuBlueleaf/KohakuRiver
    - Docker container and QEMU/KVM VM backends, VXLAN overlay network, GPU passthrough.
13. **KohakuHPO** - Unified black-box optimization system for hyperparam tuning.
    - GitHub Link: https://github.com/KohakuBlueleaf/KohakuHPO
    - S3-TuRBO and from-scratch implementation of Sobol, CMA-ES, GP-BO, HEBO, TuRBO under same interface.
14. **KohakuVault, KohakuBoard** - Key-Value/Columnar storage based on SQLite with Rust; An open sourced alternative for WandB with async logging.
    - GitHub Link: https://github.com/KohakuBlueleaf/KohakuVault
15. **(Model) Kohaku-XL** - Open sourced anime-style text-to-image generative model.
    - HuggingFace: https://huggingface.co/collections/KBlueLeaf/kohaku-xl-65e41407521cc78073e0fc0c
    - Finetuned SDXL-based text-to-image model. Have been widely used by t2i community or company.
    - Million-scale dataset (8.4M for Kohaku-XL Zeta). Trained on Consumer level hardware.
16. **(Model) TIPO and DTG** - Pretrained automatic prompt engineering model for t2i task.
    - HuggingFace: https://huggingface.co/collections/KBlueLeaf/tipo-66f9108eee826b5daad0d4bf
    - Pretrained models to be utilized in KGen library. 100M~1B LLM trained from scratch with 10~50B token seen.
    - Over 3M downloads on HuggingFace in total.

17. **(Core Dev) Stable-diffusion-webui** - Core developer and collaborator of this open-source project.
    - GitHub Link: https://github.com/AUTOMATIC1111/stable-diffusion-webui
    - Top10 contributor (#6, 124 commits, 28 merged PRs). Over 140k star
18. **(Core Dev) ComfyUI** - Core developer, tightly collaborated with ComfyUI team.
    - GitHub Link: https://github.com/Comfy-Org/ComfyUI
    - The whole weight adapter system is created and maintained by me. Also lora trainer, dataset nodes and low precision training.
19. **(Contribution) sd-scripts** - 8 merged PRs: LyCORIS integration and training fixes.
    - GitHub Link: https://github.com/kohya-ss/sd-scripts

Research Interests

- **Generative Models**
  - Small model, PEFT, Customization.
  - Image: Text-to-Image, strong control (sketch to image or AI drawing software)
  - Text: General LLM, LM for transformation task, Agent System
  - Other: Multi-modal, Voice Clone TTS, AI Singer, Sheet Music Generation
- **ML Systems**
  - Training infra, efficient kernels, AI accelerator design
- **Representation Learning**
  - Temporal, Spatial, Hierarchical
  - Real-world data
- **Ranking Systems**
  - Prediction of competition outcomes

Other Interests

- Literary, Visual Arts, Music
- Writing, Photography