Lester (Peng-Jian) Yang

AI Engineer | **LLM Systems and Evaluation**

lesterpjy@gmail.com | lesterpjy.xyz | linkedin.com/in/lesterpjy | github.com/lesterpjy | Amsterdam, Netherlands

AI engineer and graduating MSc student in Artificial Intelligence (cum laude) at the University of Amsterdam, with three years of industry experience building in-house machine-learning systems for engineering teams, across startups and global manufacturers. Merged LangChain contributions, peer-reviewed publications, and a granted US patent.

Skills

Languages and Tools: Python | LangChain | PyTorch | vLLM | OpenCV | pytest | pandas | scikit-learn | Git | Docker | SLURM | Bash/Linux

Methods: Large language models | prompting and decoding strategies | model evaluation and benchmarking | retrieval | reinforcement learning | automated testing | code review

Work Experience

Mirai Technologies, Remote | **Research Engineer (Part-time)** May 2024–Nov 2024

- Redirected Toyota's job-shop scheduling effort from better time estimates to schedule slack by showing 36% of operations ran 25% or more off plan, varying run to run on the same operation, after building a Python and pandas pipeline that turned a 3M+ row production export into schedules measured against a constraint-programming baseline.
- Cut delay passed downstream by 22% and operations knocked off plan by 20% on nine days of production data by building a slack allocator that places buffers with a bounded multidimensional knapsack over delay propagation, then proposed a Transformer and PPO policy trained against the same constraint solver, using it as both training environment and supervision signal.

Logitech, Hsinchu, Taiwan | **Senior Engineering Scientist** Jun 2022–May 2024
Engineering Scientist *May 2021–Jun 2022*

- Built a retrieval tool over the CMF engineering team's internal documents so engineers could query them directly, and contributed a PDF document loader to LangChain after the existing parser failed with no diagnostic to trace.
- Cut estimated manufacturing carbon footprint by 11%, measured from energy use, by building XGBoost models and SHAP analyses in scikit-learn that isolated the energy-intensive factors in spray painting. The reduction counted towards the annual target in Logitech's sustainability report.
- Took a precision colour-measurement system from prototype to granted US patent (US11847799B2) and into engineering verification testing, building the computer vision in Python and OpenCV.
- Built the machine-learning capability of the CMF engineering team across Hsinchu and Suzhou, turning the problems its engineers brought into systems for colour control, defect detection, and process optimisation.

Selected Projects

LangChain, open-source contribution Solo | Python, pdfplumber, pytest | PRs #4353 and #4552, merged

- Brought pdfplumber into LangChain as a PDF document loader after hitting an opaque parsing failure, chosen for its visual debugger so that a bad parse could be seen rather than guessed at. Shipped with pytest integration tests covering local and remote PDFs, plus a worked example in the project documentation.
- Rebuilt it on the parser abstraction a maintainer introduced mid-review, moving the convenience method that wrote an annotated copy out of LangChain because it did not fit the loader interface, and keeping pdfplumber as the parser so users can still reach its visual debugger directly.

Test-Time Scaling for Small Vision-Language Models Two-person team | Qwen2.5-VL, vLLM, PyTorch | arXiv:2607.09438

- Won the ImageCLEF 2026 Visual MCQ task at 84.1% on the held-out split by showing that decoding budget, not search machinery, drives test-time scaling on small vision-language models: raising the per-chain token limit from 1k to 2k gained 3.7 points where doubling the sampled chains gained 0.15, while reward-guided beam search lost to plain self-consistency at eight times the cost.

Policy Extraction in Hierarchical Offline Reinforcement Learning Solo, supervised | HIQL, PyTorch, SLURM

- Showed across 10 OGBench visual-manipulation tasks that extracting an explicit critic does not remove the known policy-extraction bottleneck, and traced the regression to approximation error introduced between the state-value and action-value functions in a data-scaling study run on SLURM.

Education

MSc Artificial Intelligence, University of Amsterdam. GPA 8.61/10, cum laude.

Master of Information and Data Science, University of California, Berkeley. GPA 3.96/4.00.

BSc Chemistry, Nagoya University

Selected Publications and Patent

- S. Baxevanakis, P. Yang. "Test-Time Scaling for Small VLMs on Multilingual Visual MCQ." *ImageCLEF* 2026. **1st place**, Visual MCQ task.
- P. Yang et al. "Color matching system for product manufacturing." US Patent US11847799B2, 2023.
- P. Yang et al. "NT5?! Training T5 to Perform Numerical Reasoning." arXiv:2104.07307, 2021.
- P. Yang, M. Sugiyama, K. Tsuda, T. Yanai. "Artificial neural networks applied as molecular wave function solvers." *Journal of Chemical Theory and Computation* (ACS), 2020.