Yen-Nan (Aaron) Ho

New Taipei City, Taiwan | Website: https://aaron1aaron2.github.io/
Email: hoyennan@gmail.com
AI Engineer

Professional Summary

Data Scientist at Harbor Tech focused on industrial AI products and customer delivery. Lead product/domain design for Harbor AI 2.0 monitoring and an equipment inspection project from requirements through on-site delivery. Combine data analysis, system integration, and AI-assisted engineering to turn diagnostic methods into usable, verifiable products.

MCIENet, my M.S. thesis, is my most complete research case, spanning models, data pipelines, ablations, and cross-dataset evaluation. Other work includes passenger-flow estimation, financial text analysis, and person ReID.

Technical Skills

- **Data Science and Industrial Diagnostics:** Python, NumPy, pandas, SciPy, scikit-learn, PyTorch; vibration/PRPD, FFT/WAV, RUL, statistical tests/effect sizes, CNNs, ablations, XAI, and data-availability assessment.
- **AI and Reporting Systems:** LangChain/LangGraph, vLLM, OpenAI-compatible APIs, deterministic analysis separated from LLM narrative, per-device isolation, fallbacks, and human validation.
- **Backend and Product:** Django/Django Ninja, Pydantic, PostgreSQL, Redis, RabbitMQ, REST, asynchronous workflows, React/TypeScript.
- **Engineering and Delivery:** Docker/Compose, offline/cross-architecture installation, update/rollback, migration, Git, automated tests, AI agents, and human release review.

Work Experience

Data Scientist, Harbor Tech | Taipei, Taiwan | Oct 2025-Present

Worked on two Harbor AI 2.0 products sharing a technical foundation: ongoing equipment monitoring/diagnostics and unit-by-unit end-of-line quality inspection. I lead product/domain and customer-delivery work; AI agents and engineering partners contribute implementation and verification as described below.

Industrial Intelligent Monitoring Platform: From Domain Workflows to an Independent Product

- **Product and operator workflows:** Led product/domain design and built much of 2.0's functionality from scratch with the team. Translated vibration, PRPD, RUL, trends, and data quality into layered equipment/sensor status, event views, and operator workflows. Used a device-first reporting flow and connected multilingual interfaces, charts, mobile layouts, and live APIs.
- **Legacy-system evolution:** Defined ownership of equipment data versus user settings during coexistence, protecting new settings from legacy synchronization. Advanced device-control transfer, event-source replacement, and auditable migration; stopped imports when sensor mappings were unclear and revalidated after corrections. Confirmed in October 2026 that 2.0 operates independently of 1.0; this does not imply every customer or dataset has migrated.
- **Analysis workflows:** Collaborated with platform engineers on Django orchestration, defining APIs, data access, artifact-by-reference contracts, task lifecycles, and failure behavior. Integrated DB-authoritative per-sensor scheduling, manual triggers, batches, and history, choosing aggregate or WAV inputs by availability.
- **Edge data and control:** Owned Gateway ingestion specifications and design validation across uploads, heartbeats, duplication, integrity, and capacity, handing server implementation to the responsible

engineer. Also contributed server-side software-package distribution and RUL-to-threshold feedback across platform state/history and Gateway updates, keeping ingestion, migration, OTA, and threshold policies distinct.

Equipment End-of-Line Quality Inspection Platform: Requirements, Operation, and On-Site Delivery

- **End-to-end responsibility:** Led customer requirements, specifications, implementation, UI/UX improvement, and on-site delivery with agents and engineering partners. Defined unit/batch relationships, reference samples, multipoint measurement, interrupted orders, serial/work-order continuity, human review, and reports. Kept functional verification separate from on-site decision-threshold calibration.
- **Operator experience:** The initial AI-generated UI met specifications but required five to six page levels and circular navigation. Reorganized daily work into plan-list, work-order, and board interfaces, integrating reports, review, and reference management or moving them into dialogs. Added focused primary actions, arrow/Enter controls, status cues, and sensor-disconnection reminders. DQA feedback was qualitative.
- **Measurement safeguards:** Separated sensor-installation confirmation from measurement, using sensor IDs and installation diagrams as mandatory pre-measurement checks.
- **On-site judgment and fixes:** Installed/demonstrated the system, trained users, and answered customer questions on October 12, 2026. Personally diagnosed a clock offset behind interrupted connections, executed the troubleshooting commands, and completed on-site and follow-up fixes. Claude supported discussion of symptoms and hypotheses. Clarified feedback for the team and explained operation and design choices to users.
- **Delivery status:** Coordinated DQA, test versions, manuals, and acceptance records. The project remains ongoing; formal sign-off and final acceptance remain unconfirmed. My estimate of roughly 80% overall contribution covers the end-to-end work, not a measured code-authorship share.

Models, Diagnostics, and Reports: Keeping Conclusions Tied to Data

- **Assumption review:** Traced RUL analysis to test constants treated as observations, rejected the conclusion, and required multi-parameter/multi-curve and forward/reverse ablations. Connected research notebooks to product code to check rapid-degradation misses, healthy-series false alarms, and implementation drift.
- **Diagnostic/output rules:** Connected preprocessing, valid operating segments, fault classification, ISO severity, RUL, and reporting while keeping fault type, severity, and remaining time distinct. Reviewed sampling, coverage, selection bias, and denominators; explicitly degraded unavailable results and preserved engine outputs. Field lifetime ground truth and confidence calibration remain incomplete, so I do not claim RUL accuracy.
- **Report-service productization:** Turned manually validated field reports into reusable methods and quality rules. Used a July LangChain/LangGraph POC to validate node contracts, device isolation, partial regeneration, and fallbacks. Deterministic nodes produced values/classes; LLMs organized narrative. Advanced service/platform integration in August and mainline DOCX/frontend downloads in September. Led methods and quality criteria with team/agent implementation and integration; deployment and customer acceptance are tracked per path.
- **Model-service interfaces:** Validated live vLLM chat, embedding, and tool-calling capabilities, separating connectivity, functional, and semantic health checks. Endpoint availability is not evidence that every report path has switched providers.

AI-Assisted Engineering and Release Delivery: Making Parallel Work Traceable

- Established shared requirements, analysis, decisions, issues, tests, and version records. Decomposed work, supplied persistent context/tools, and set verification/stop criteria; agents researched, implemented, repaired, and rechecked changes on isolated branches.
- Ran daily functional regression and monitoring, routing findings to issues and Bitbucket PRs. Reviewed data assumptions and product behavior with human reviewers before merge/release. Fed recurring problems into reusable rules and tests; no overall productivity multiplier or automatic-repair success rate has been measured.

- Integrated Docker/offline packages, cross-architecture installation, version information, and platform update/rollback. Separated update execution from platform downtime risk, with prechecks, progress, failure signals, and recovery. Used real-machine failures to improve write freezes, schema/constraint, storage, and restart checks.
- Refined delivery-package contents, excluded customer raw data, validated clean installations, and led release integration/readiness decisions. Migration checks covered fields, time zones, readability, and continuity; linked code, issues, versions, and manuals through VitePress.

Field Analysis, Technical Delivery, and Collaboration

- Aligned field/quality teams on DB, event, and WAV scope for semiconductor and large multi-sensor partial-discharge analysis. Used the full database to locate events, then retrieved corresponding waveforms for batch analysis, accumulated PRPD images, AI drafts, and page-by-page human review.
- Used max/P99 to distinguish sustained motor/pump severity from one-second spikes, checking operating segments, RUL basis, and VRMS variability. Later used minute medians, nonparametric tests, multiple-comparison correction, and effect sizes to reassess customer observations. Delivered the analysis and customer explanation; root cause still requires field speed, flow, and related information.
- Reconstructed legacy code, notebooks, and reports into traceable ADC-to-g, FFT, SNR, cutting-point, and frequency-domain workflows, delivering executables/manuals while protecting company algorithm IP.
- Designed offline research topics, learning material, and starter notebooks, reviewing data assumptions/direction through weekly reports; the intern retained authorship of their analysis. Also owned company-side technical drafting, version convergence, and interface/checkpoint alignment for a smart-city edge-platform proposal, which remained an unsubmitted draft.

Data Science Intern, Institute of Sociology, Academia Sinica | Taipei, Taiwan | May 2019-Dec 2020

- Combined official sample addresses, point-to-point/fallback distances, survey-area grouping, OR-Tools TSP, and Folium for TSCS field routes, dispersion allowances, and budget estimates.
- Used Selenium, multiprocessing, Tor multi-port routing, and Folium for distance collection and GIS outputs as the project's contemporary technical approach.

Representative Research and Projects

MCIENet | M.S. Thesis, 2024

- Predicted 3D chromatin interactions/CTCF loops from DNA sequences using multi-scale CNN/Inception architecture, examining signals from different receptive fields.
- Independently led research decomposition, model and data-pipeline integration, experiment design/execution, ablations, cross-dataset evaluation, XAI, thesis writing, and reproducible repository engineering.
- Used BaseCNN comparisons, sequence-length settings, and ablation configurations to examine design choices. Retained DeepLIFT, ChINN/DNABERT2 comparison entry points, and shared split configurations for traceability. Findings are limited to completed experiments; comparison entry points and shared configurations were retained for further reproduction.
- Treated dataset, positional, and resolution differences as generalization checks, retaining cross-cell-line limitations. Adapted parts of positive/negative preprocessing from ChINN and distinguished that prior work from my model, experimental, and engineering contributions.

Financial-Text Sentiment Analysis | TCSE Research, 2020

- Studied stock-article sentiment using a Hierarchical Attention Network and sentiment lexicons. The work began with an approved undergraduate research grant and led to a TCSE publication.
- Performed all research work for the study; the paper was coauthored, so this scope does not imply sole authorship.
- Early repeated model changes yielded limited improvement. Gradually added data and parameter comparisons, retaining comparison conditions and experimental results as the basis for decisions.

Dynamic Embedding for Passenger-Flow Estimation | IIAI AAI Research, 2021

- Explored dynamic embedding for passenger-flow estimation. Responsible for method ideation, data processing, model construction, and experiment execution.
- Used data and parameter comparisons to examine model choices. The research received an IIAI AAI Competitive Paper Award; publication credit remains shared.

Research Experiment Practices Developed Over Time | 2020-2024

- Progressed from repeated model changes with limited improvement toward data and parameter comparisons, then more complete ablation design and parameter-adjustment logic in the MCIENet thesis.
- Used Shell to organize parallel experiments under limited GPU resources, including running two parameter configurations on one GPU with separate logs and checkpoints.
- Saved preprocessed data as H5, used early stopping, and first sampled three parameter values to inspect the curve before deciding how to adjust the search.
- Developed these practices across the 2020 sentiment study, 2021 passenger-flow study, PropGMAN, and MCIENet, with different levels of use in each project. This does not imply an identical complete pipeline across all projects or a measured overall speedup.

Dynamic Traffic Network Representation | KBS Collaborative Research, 2024

- Coauthored research on dynamic traffic network representation for passenger-flow prediction. My actual contribution was primarily supporting code and technical assistance.

Horus Human Identification System | Undergraduate Capstone, 2020

- The team integrated multi-object tracking, person re-identification, and face recognition into a multi-camera identification and tracking system.
- Primarily owned ReID feature-extraction integration, folder sampling/matching, cosine similarity, dual-threshold decisions, and tracking-record queue orchestration.
- Integrated the third-party FastReID base model with the team’ s MOT and face-recognition modules. The team won first place in Soochow University's undergraduate capstone competition.

PropGMAN Asset Activation | Team Project, 2022

- In a three-person team, applied spatiotemporal modeling to property-transaction records to predict regional land prices and transaction volumes for asset-activation analysis.
- Proposed the project direction, led team coordination and division of work, and handled model implementation, experiments, and the final-round report. Retained preprocessing, model settings, experimental results, and reproduction instructions. The base GMAN method used an existing third-party research implementation. The team received a Creative Award in the AI for Asset Activation Innovation Contest.

Additional Independent Projects

Dynamic Bus Route and Schedule Planning | Project period to be confirmed

- Integrated tourism demand, Google Trends, Popular Times, attraction locations, and travel distances for route and schedule planning in eastern Taiwan. Built demand-allocation and distance-matrix workflows, explored OR-Tools TSP routes, and visualized routes and schedules with Folium, Gantt, and Sankey views.
- This was an exploratory independent project, separate from the Academia Sinica survey-routing work. The project period remains unconfirmed; no transit-operator adoption or operational outcome is claimed.

Additional Team Projects

- **Distributed Spark Crawler/Recovery, 2022:** An NCCU distributed-systems course team connected collection, missing-data detection, Spark-based imputation, a database, and dashboards to demonstrate recovery from collection failures. I owned the Spark cluster, GitHub work, and demo video, and contributed system integration.
- **Financial Recommendation and Automated Notifications (E.SUN Fugle), 2020:** Proposed the topic direction and led team coordination and division of work. Primarily designed, implemented, and deployed the Telegram chatbot, integrating modules from teammates into one entry point. The team used item-based, user-based, and co-occurrence recommendation approaches; I offered algorithm ideas and suggestions rather than implementing every algorithm myself. The team received an Excellence Award in a financial-programming course competition.
- **Taipei Disaster Resource Dashboard, 2025:** Proposed the topic and collected data for a team analysis of disaster-prevention resources around tourist attractions. The team reached the Taipei-New Taipei City Dashboard Hackathon finals. The public dashboard is based on a fork of a friend' s template, with team and template credit retained.
- **TownPass Authorized Assistance Proposal (2026):** Owned topic ideation, use-case definition, and product specifications, and presented in the final round. The team reached the finals of the 2026 TownPass Microservices Hackathon. Specifications covered delegation scope, user flows, data permissions, API contracts, and acceptance criteria; the final presentation described a single-case simulated workflow requiring the applicant' s confirmation.

Research and Publications

- Dynamic traffic network representation model for improving the prediction performance of passenger flow. Knowledge-Based Systems, 2024. Primarily contributed supporting code and technical assistance.
- A Dynamic Embedding Method for Passenger Flow Estimation. IIAI AAI, 2021. Responsible for method ideation, data processing, model construction, and experiments. Competitive Paper Award.
- Hierarchical Attention Network with Sentiment Lexicons for Stock Article Emotion Analysis. TCSE, 2020. Performed all research work; the paper was coauthored. Preceded by an approved undergraduate research grant.
- MCIENet: Multi-scale CNN-based Information Extraction from DNA Sequences for 3D Chromatin Interactions Prediction. M.S. thesis, 2024.

Education

- National Chengchi University, M.S. in Computer Science | Sep 2021-Sep 2024.
- Soochow University, B.S. in Data Science | Sep 2017-Jul 2021.

Recognition

- IIAI AAI Competitive Paper Award, 2021.
- Creative Award, AI for Asset Activation Innovation Contest, 2022.
- First Place, Soochow University Undergraduate Capstone, Horus, 2020.
- Excellence Award, Financial Programming Course Competition, 2020.
- Finalist, Taipei-New Taipei City Dashboard Hackathon, 2025.