

ฮิโระ ทานากะ

วิศวกร AI (อาวุโส)

ข้อมูลติดต่อ

อีเมล: hiro.tanaka@example.com

โทรศัพท์: +81 70-1234-5678

ที่อยู่: โตเกียว, ญี่ปุ่น

<https://linkedin.com/in/hirotanaka> |

<https://github.com/htanaka> |

<https://hiroai.dev>

วันเกิด: 1988-02-04

สัญชาติ: ญี่ปุ่น

สรุปโปรไฟล์

วิศวกร AI อาวุโสที่เชี่ยวชาญในระบบ LLM

การประเมินผล

และการเพิ่มประสิทธิภาพด้านความหน่วงและต้นทุนใน
ระดับใหญ่

ทักษะ

- PyTorch
- การเทรนแบบกระจาย
- การเพิ่มประสิทธิภาพ
- RAG
- วิศวกรรม prompts
- การประเมินผล
- MLOps

ประสบการณ์ทำงาน

วิศวกร AI อาวุโส — AI Corp Japan (โตเกียว, ญี่ปุ่น)

มี.ค. 2020 – ปัจจุบัน

- นำทีมพัฒนาและปรับใช้ LLMs ในผลิตภัณฑ์หลักขององค์กร
- ลดต้นทุนการอนุมาน LLM ลง 60% ผ่านการเพิ่มประสิทธิภาพ

เทคโนโลยี: PyTorch, TensorRT, ONNX, Kubernetes

วิศวกรแมชชีนเลิร์นนิง — Cyberdyne Systems (โอซาก้า, ญี่ปุ่น)

เม.ย. 2015 – ก.พ. 2020

- พัฒนาอัลกอริทึมการเรียนรู้เชิงลึกสำหรับระบบหุ่นยนต์
- ปรับใช้โมเดล ML ในสภาพแวดล้อม edge และ cloud

เทคโนโลยี: Python, TensorFlow, ROS

โปรเจกต์

LLM Optimization Framework — Python, PyTorch, ONNX

พัฒนาเฟรมเวิร์กสำหรับเพิ่มประสิทธิภาพ LLMs แบบโอเพนซอร์ส

<https://github.com/htanaka/llm-opt>

มี.ค. 2022 – ปัจจุบัน

การศึกษา

ปริญญาเอก — มหาวิทยาลัยโตเกียว

คณะ: วิทยาการคอมพิวเตอร์ • สาขา: การเรียนรู้ของเครื่อง
2010 – 2015

ใบรับรอง

- AWS Machine Learning – Specialty

รางวัล

รางวัลนวัตกรรมดีเด่น — AI Corp Japan, 2022

สำหรับการพัฒนาเทคนิคการเพิ่มประสิทธิภาพ LLM

ภาษา

- ภาษาญี่ปุ่น — ภาษาแม่
- ภาษาอังกฤษ — คล่องแคล่ว