

Mohamed Khodeir

m.khodeir@mail.utoronto.ca • +1 (647)-224-9690

EDUCATION

- MSc in Computer Science, University of Toronto** SEP 2021 - MAR 2023
Research focused on the interplay of planning and machine learning in Robotics. Supervised by Prof Florian Shkurti at the [Robot Vision and Learning Lab](#).
- BSc in Computer Science, University of Toronto** SEP 2011 - APR 2015
Relevant Coursework Machine Learning, Natural Language Computing, Parallel Programming

WORK EXPERIENCE

- Machine Learning Engineer at Xero** JUN 2019 - SEP 2021
Generalized Financial Document Extraction API
- Developed an auto-scaleable state of the art document understanding service used by multiple Xero products, backed by SQS, running in kubernetes and maintaining stringent SLAs while processing a peak load of ~100K documents per hour.
 - Led development of a highly fault-tolerant service to extract text, layout metadata and images from popular file formats (e.g. pdf, html, jpeg, doc).
 - Trained, evaluated and deployed transformer-based named entity recognition models. Experimented with approaches for improving robustness to sensible re-ordering of input text.
 - Designed and implemented a Python library for defining parallel multi-step data pipelines. Used this to 3x GPU utilization during multi-GPU training jobs.
- Data Scientist at Hubdoc** OCT 2017 - MAY 2019
Human In The Loop Document Extraction
- Implemented, trained and deployed machine learning models for named entity recognition and document classification. Helped design a methodology for tuning confidence thresholds on noisy labeled-data.
 - Worked with the team to drive recall of predictions from 30% to 80% over the course of a year, while holding accuracy at 90%. This meant users got their data in seconds instead of hours while the company lowered human labelling costs by 50%.
 - Implemented a queue-based model server to enable predictions at scale. Instrumented this with performance and quality metrics that incorporated delayed user feedback.
 - Co-led a cross-team development/deployment which refactored a synchronous API monolith into an async API with a modular queue-worker pipeline architecture. This was necessary for improving scaling properties of the service and allowing integration of arbitrary background processing including OCR.
- Co-Founder at Tutorama** JAN 2016 - OCT 2017
Curated Tutoring Marketplace
- Bootstrapped ~50,000USD from placing 1st in regional startup competitions including MIT-EF and Seedstars.
 - Developed a web app with React for facilitating a marketplace interaction between students and tutors. Managed thousands of hours of tutoring per month at peak.
 - Implemented API-level integrations for credit card payments, email/sms notifications

PUBLICATIONS

- Policy-Guided Lazy Search with Feedback for Task and Motion Planning** SEP 2021 - DEC 2022
- Led research in developing a planning algorithm for task and motion planning which can be guided by a learned policy. Received the [best paper award at a CoRL workshop](#) for long-horizon planning. Publication to appear in the IEEE International Conference on Robotics and Automation (ICRA) 2023. [Preprint](#).
- Learning To Search in Task and Motion Planning with Streams** JAN 2021 - DEC 2021
- Led research in learning Graph Neural Networks as heuristic functions for task and motion planning problems in robotic manipulation. [Published in IEEE Robotics and Automation Letters](#).
- Evaluating robot task planning over large 3D scene graphs** SUMMER 2021
- Co-authored a large-scale benchmark of planning approaches applied to 3d Scene Graphs. Publication appeared in the Conference on Robot Learning (CoRL) 2021. [Project Page](#).

PROJECTS

- AI for Contract Review | [Legalink.tech](#)** MAY 2020 - JAN 2022
- Developed NLP-based recommendations for an MS Word Add-In for editing legal contracts.