Ahmad Badary (AJ)

504-1320 1 st se, Vancouver, AB T2G 0G8   ✳   (403) 608-2592   ✳   badary.ahmad@gmail.com   ✳   https://ahmedbadary.github.io

Summary

Strong background in Machine/Deep Learning, Statistical Learning Theory, and Optimization. Expertise in Natural Language Processing/Understanding problems, in particular, and other AI applications in general. Research experience in Deep Learning, Learning Theory, Quantum Computing, Particle Physics, and Information Theory (compression). Specific work on generative models, generative compression (image/text), embeddings, medical NLP, natural language dialog systems and assistants, and knowledge graphs.

CORE COMPETENCIES

- Machine/Deep Learning
- Software Engineering
- Leadership & Managerial Experience

- NLP/NLU
- Theoretical/Applied Quantitative Skills
- Communication & Presentation Skills

Academics

UC BERKELEY (2014-2017)

- *Major(s)*: Computer Science, Applied Math, Physics[*]
- *Relevant Coursework:*
    - CS 231n – Convolutional Neural Networks for Visual Recognition [image classification, localization, and detection.]
    - CS 224n – Natural Language Processing with Deep Learning [Sequence Modeling, Recurrent Nets, NLP] (Stanford)
    - CS 289 – Machine Learning [Regression, Decision Trees, Ensemble Learning, Neural Networks, SVM, Clustering]
    - EE 227 – Optimization Models [Convex / Conic Optimization, Duality, robustness, stochastic programming]

Work Experience

- **Tofu AI [11/2022-]**
    - **Title:** CTO (Co-founder)
    - **Desc:** Co-founder and CTO at Tofu AI: Fully Automatic BDR and Sales Playbooks Framework Generation. Tofu AI leverages SOTA AI models to help companies empower their Business Development team to build better pipelines faster with the right messaging and positioning. Tofu AI is an AI start-up that caters to businesses looking to automate their BDR process. I built the backend and AI models, pipeline, and infrastructure for our software. Models built include Large Language Models (LLMs) based on Transformer architecture for automatic generation of personalized communication material (emails, messages, call scripts) for prospective customers with strong focus on maximizing customer acquisition. Models also automatically process the information about the products and services offered by our clients and integrate that information seamlessly into the generated communications. More info at: tofuai.co

- **Google AI [05/2022-]**
    - **Title:** Machine Learning Researcher (Consultant/Fellow)
    - **Desc:** Collaborator on various research projects in ML/DL focusing on Large Language Models (LLMs) and Transformer Networks. Projects and papers include *Linearized self-Attention, Non-AutoRegressive Language Modeling, Maximum Data Compression lossy/lossless (Text), Emergent Abilities, AGI.*

- **Time by Ping (~Laurel AI) [05/2020 - 05/2022]**
    - **Title:** Director, AI
    - **Desc:** : Led the AI team at Time By Ping, an Enterprise SW company utilizing AI in the Law space. Problems included capturing lawyers' work on various devices and classifying (assigning) it to the different clients and cases they were working on, as well as Grouping relevant "units" of work together. Anonymization of the data collected. Building a stable, resilient AI infrastructure and data pipeline. Adding natural language descriptions to the units of work using advanced NLP techniques. Finally, matching different units of work to their billable codes to speed up the billing process.

- **Mobsquad Inc. [04/2019 - 05/2020]**
  - **Title:** Machine Learning Software Engineer
  - **Desc:** Consulted for multiple clients as an Applied Research Scientist on problems of Text Classification, Data Anonymization, Activity Segmentation, Summarization, Topic Modeling, Billing-Codes Assignment/Classification. Helped multiple AI departments on core areas in NLP and NLU.
- **SUKI AI [05/2018 - 02/2019]**
  - **Title:** AI/ML Tech Research Engineer - NLP Focus
  - **Desc:** The first AI research engineer at Suki AI. I worked on many NLP/NLU problems in the personal-assistant space. Designed+Built a complete Intent-Extraction system and a Slot-Extraction/Filling system. Built a complete Medical-Model Ontology (unsupervised). Designed+Built a complete End-to-End "Online-Learning Model Management System". Collaborated with Amazon AWS ML team and Google AI team to help build specialized ASR (Speech Recognition) models catering to the Medical setting.

Qualifications and SKILLS

- **PROGRAMMING**
  - Python [Proficient] + (Flask - MongoDB - SQL)
  - GoLang [Proficient]
  - Java / C / C++ [Familiar]
- **Specialized Skill-Set**
  - Machine Learning / Deep Learning
  - AI / Autonomous systems
  - TensorFlow / PyTorch / Keras / JAX
  - Natural Language Processing: LLMs, Transformers, RNNs (LSTMs, GRUs), AutoRegressive Models
  - Analytic Algorithms [Page-Rank, Streaming, Sampling, Mult.weigh update, Boosting, randomized algos]
  - Graph Theory / Spectral Graph Theory / Graph Clustering and Partitioning

Research and Projects

- **Computer Science and Tech**
  - State of the Art: Generative Text Compression [Project + Paper (pending)] (UC Berkeley)
    Built SOTA text compression using intelligently learned transformations of the input to reduce the overall entropy of the text.
  - State of the Art: Intent Extraction/Parsing System [Paper+Project] (Suki AI)
    Built SOTA system for extracting intent from natural text - part of an intelligent assistant agent. 99.7% Accuracy / 0.98 F1. 41% relative improvement over Google's DialogueFlow and 48% over Amazon's Intent Parser. The system/model was in place for 4+ years without any revisions.
  - State of the Art: Slot Extraction/Filling System [Paper+Project] (Suki AI)
    Built SOTA system for Slot Extraction and Filling from natural text - part of an intelligent assistant agent. 98.4% Accuracy / 0.97 F1. 40% relative improvement over Google's DialogueFlow and 46% over Amazon's Intent Parser. The system/model was in place for 4+ years without any revisions.
  - 2 Full Patents and 2 Provisional Patents on Voice-Based Assistant Technology [Patents] (Suki AI)
    Multiple Patents created for Suki AI spanning various NLP/NLU technologies including Intent Parsing, Slot Filling, Text Embeddings (word/sentence/document-levels), as well as an ingenious method to improve on the theoretical limit of accuracy of the voice commands by amending the ASRs outputs provided by Google's Speech Recognition system to be more geared towards medical vocabulary.
  - SOTA Model for Dynamic Multiclass Classification of "Work Units" to intended Clients [Project] (TBP - Time By Ping AI)
    Built a model that can classify "chunks" of work performed by users (lawyers) captured on their devices (computers, phones, etc) into their respective clients for tracking work and billable hours. The users' clients were dynamically changing as old clients leave and new clients come on board.

- o Brain-Controlled Vehicles [UC Berkeley]

  Signal processing. Working with EEG waves and analyzing how they are used to control and manipulate certain vehicles, i.e. quad-copters. Succeeded in refining a navigation system by using SOTA deep learning to model the EEG waves produced by the brain.

- o The VR-y Night [Cal-Hacks]
  Link: https://devpost.com/software/stylistic-rendering-of-street-view-with-motion-control
  "Our project is a VR environment that uses deep learning techniques to transfer artistic style to any given panorama. We can then visualize the panorama with Oculus, and with Google street view we can do it for any real-world place, making it feel like the person is walking around in a painted world."

- o Wind Turbines Optimization [Patent]
  A patent for a system based on LIDAR technology, that uses gradient optimization methods and machine learning to analyze the dynamics of the wind and optimize the rotational directions of mechanical wind turbines to optimize the global (overall) energy converted by a wind farm.

- **Physics**
  - o Assessing the Potential of the Casimir Effect. Looking into potential uses of the Casimir effect in, clean renewable energy-based, jet propulsion systems.
  - o Methods for a Mathematical Formulation of Quantum Mechanics in 5-D [Cairo, Egypt]

    Developing mathematical tools, 5-d metric space and a working field, for a framework of modern physics working in four-dimensional Spacetime plus one spatial dimension in the hope of resolving the discrepancy of wave-particle duality.

  - o Ongoing Mathematical Study of the Geometry of Spacetime in General Relativity [Independent]

    An attempt at a more rigorous formulation of the Geometry of Spacetime using Minkowski spaces and the Lorentz group. Ref: Group Theory and General Relativity|Carmeli / Comprehensive Differential Geometry|Spivak.

Teaching Experience

- **Computer Science 289 | Machine Learning [UC Berkeley]:** TA (Academic Staff)
- **Computer Science 61C | Computer Architecture [UC Berkeley]:** Lab TA (Course Staff)
- **CS 61a/61b/170/189/70 + Math 1a/1b/113/110/104/202a/b [UC Berkeley]:** Tutor

Patents

- **Systems And Methods For Generating Disambiguated Terms In Automatically Generated Transcriptions Including Instructions Within A Particular Knowledge Domain**
  US-11217227 · Issued Dec 15, 2021
- **Systems And Methods To Facilitate Intent Determination Of A Command By Grouping Terms Based On Context**
  US-11538565 · Issued Dec 07, 2022
- **Yaw Fitting Using Gradient Descent on Wake Centerline Points**
  US-11047362 · Issued Jun 29, 2021

Writings and Publications

- **[Refer to Personal Website (above)]**