

## Summary

As collecting large amounts of data becomes the norm rather than the exception, learning to deal with the storage, processing and analysis of this data is becoming a looming problem in industry. Industry is often lagging behind academia in terms of internalizing new ideas and advancements in this area. Every company I have worked for has been struggling to store the large amounts of data that their users have been generating efficiently and accessibly, and to utilize this mass of data to make their product more intelligent through machine learning methods such as classification or recommender systems.

In light of this trend, I would like to pursue an M.S. in Computer Science, with a focus on databases and machine learning. I hope to spend an intensive period of study by taking courses and doing research in order to deepen my knowledge and improve my skills in these areas.

## Research Experience & Interests

### Undergraduate

In my senior year at WPI, my curiosity toward machine intelligence led me to Prometheus, a project to build an autonomous robot that could traverse an obstacle course and navigate GPS waypoints. The goal was to have the robot participate in the [IGVC](#) (Intelligent Ground Vehicle Competition) and the ION (Institute of Navigation) [Lawnmower competition](#).

I worked as one of two Computer Science students in a six person interdisciplinary team led by Profs. [Taskin Padiir](#) and [Stephen Nestinger](#). The project involved taking a robot that was already physically built for the previous year's IGVC and adding intelligence capabilities for autonomous navigation in an obstacle course. It continued throughout the entire school year and into the summer.

My work consisted of several tasks. The first and most important was to devise and implement lane detection functionality. I utilized the dual cameras on the robot to accurately detect the location of lane lines that the robot should not touch, and placed them on a local probability map of obstacles, along with data from other sources like LIDAR (laser range detection). I abandoned a GPU processing based approach (with CUDA) in the prototype stage in favor of a simpler OpenCV based solution, since line detection was not a performance bottleneck. This allowed me to simultaneously help my team on a variety of other tasks, including:

- Writing code to process realtime GPS data, and perform geodetic and coordinate system conversions to be able to localize the robot relative to GPS waypoints.
- Setting up the general architecture of the robot (using [ROS](#)). This included a networked pub/sub system between different data sources, processing nodes and computers. Sensor nodes streamed data to processing nodes, which did the requisite calculations and fed them to actuators.
- Aiding in developing other intelligence-related functionality such as the probability map, SLAM integration, etc.

- Aiding in various debugging, evaluation, testing, integration and prototyping tasks with the robot and with new hardware and sensors.

The project was published in the [SPIE conference proceedings](#). My method and implementation of the lane detection was robust enough that the following year's team opted to incorporate it unchanged.

In addition, while I was working on Prometheus, I discovered possible improvements by applying what I had learned in my senior year AI course. The lane detection rarely failed, but had to be calibrated every time the robot was brought onto the field. This was because each scenario had different lighting conditions (sunlight vs fluorescent) as well as different shades of paint used for the lines and different colored ground.

It occurred to me that instead of tweaking parameters by hand until I obtained good results, a method such as hill climbing or simulated annealing could accomplish the same task automatically. Also, lane markers of a specific color form a pattern that is simple enough to express in terms of basic lines, but more complicated obstacles such as cones could be better detected by using ML classifiers such as Haar, SURF and SIFT. This would be better than being detected by the laser scanner and not being distinguished from a wall or other obstacle.

Although these improvements were left at the prototype or idea stage because of the severe lack of time, they rekindled my interest in artificial intelligence. This motivated me to simultaneously take [a graduate level course on autonomous multi-robot systems](#) taught by Prof [Stephen Nestinger](#) that explored cooperative robot intelligence problems, and thus had a distributed systems focus.

## Industry

After graduation, I found myself facing similar problems in a different context. At my own startup, Bitwich LLC, I faced the challenge of taking multiple messy and non-uniform raw CSV datasets from the government, and unifying their common aspects into one large dataset. Part of the results of my effort is now [open sourced as a Python library called transformer](#) that allows one to very easily extract a few fields and apply arbitrary transformations such as merging with another field, regex transformations, numeric conversions, etc. Transformer takes this normally tedious and repetitive task full of edge cases and reduces it to a few minutes of simple programming.

This solved half of the data unification problem, but the other half remained a manual labor task: How does one look at two data fields that are determined to be related, and automatically transform them into a unified format? For example, a date in one dataset could be in year-first YYYYMMDD format, while the other could be standard MMDDYYYY, and include time. Randomly sampling said data and applying classification algorithms to determine the format would make the entire process automatic. Relating multiple different data sources unlocks information that was previously unavailable by looking at each individual source. Automating the process of determining data structure via machine learning could be immensely valuable.

Recently, I brushed with machine learning yet again while working for [CustomMade](#), an online marketplace. I joined amidst an effort to make the marketplace more intelligent by

leveraging user and product data. The site connects artisans with people looking to have custom goods made. Customers post a text description of their project idea and attach photos. We are experimenting with using a classifier to be able to accurately place an incoming post onto a taxonomy of product types. Furthermore, we are trying to detect posts that are unlikely to be selected by our makers, by training a model on features of those posts that have successfully filled including content, specific words used, time of post, whether an image was attached, and other similar factors. Even a simple boost in any of these areas is very profitable.

Another project I completed at CustomMade involved architecting a system for storing large amounts of user to user and system to user message data, including transport-specific information such as e-mail and SMS data. The project involved a sharded database setup that split the data into separate metadata and storage portions based on empirically measured access patterns. It turned out that storing all the data together required compromising one kind of query for another (range scans and aggregate queries vs random access). However, partitioning the data according to the kinds of queries that would fetch it proved to yield a high-performance solution.

All of the aforementioned encounters motivate me to learn more about methods relating to Databases and Machine Learning. I hope to enter the WPI M.S. CS program, take courses relating to these issues while also engaging in research, and to leave with a deeper understanding therein.

## Possible Research

I humbly suggest that I would make a good research assistant. I know it is less common for master's students to engage in research compared to PhDs, but I hope I've adequately demonstrated a passion for my area of interest, my eagerness to learn quickly and my ability to execute. Additionally, my industry experience means that I can implement your research projects fast and well, and in a manner that is easily reproducible by your colleagues. This is task that I know some professors view as menial, but I disagree. I enjoy this as much as the formal research, and I would be enthusiastic to help in this regard also.

On the ML side, Prof. Chernova's work on online Learning from Demonstration fits my interests well. I imagine my experience with Prometheus is a good starting point for the robotics portion of the project, and nearly the entirety of my work experience is related to the web application portion. This would be a great opportunity for me to learn more about classification, planning and exploration algorithms, all under the umbrella of supervised robot learning.

I would also be interested in working with Prof Ruiz, based on her previous work on recommender systems, such as her work on using neural expert networks for recommendations and her earlier work on association rule mining for recommender systems. Her applications work, such as on machine learning clinical performance in a pancreatic cancer database or modelling student performance for predictions also seem relevant.

On the Databases side, I would be excited to work with Prof. Eltabakh. His work on Hadoop, including Eagle Eyed Elephant, CoHadoop and JAQL, and his work adding SP-GiST to PostgreSQL is exactly the kind of information I would like to accumulate. My experiences working hands-on with large amounts of data both on data warehouses like Hive + Hadoop

and conventional databases like MySQL and PostgreSQL led me to often think about issues similar to those presented in his papers.

I am open to working on other research projects that I am yet unaware of, which involve similar categories of problems involving machine learning and databases.

## Assistantship

I am also interested in taking on work as a teaching or research assistant. I have teaching experience, ranging from tutoring my own friends throughout high school and college to teaching in formal settings. In my junior year, I served as a teaching assistant for Prof. [Kathi Fisler](#) in what was then WPI's experimental [Software Security course](#), which has expanded into [its own sub-department](#) since.

After graduating, I've taken an active role in the [Boston Python group](#), which is the largest Python group in the world and largest programming related group in Boston. I've participated in many project nights where I've helped beginners reach milestones in their projects. I've also talked in front of more than 100 people at two of the presentation nights, once [explaining the thought process](#) behind a [distributed ID generation library I wrote](#), and another time doing a [live-programming](#) demonstration in front of the audience to demonstrate the user friendliness of a boilerplate template I developed for creating REST services.

The experience I've gained has given me the skills to organize my thoughts, communicate effectively with an active audience, and most importantly transfer knowledge, all of which would be useful in a teaching assistant position.

## Other Experience

Working in industry, I have also gathered experience that is relevant to doing research and obtaining a graduate degree. Many of the projects I worked on were in analogous circumstances to that of academic research: They usually started out with not much more than an idea that could lead to innovation, had minimal formal guidance, meant considerable work beyond the average 9 to 5 routine, required rigorous technical thinking, and were relentlessly focused on bringing in results.

My experiences are as varied as adding fingerprint authentication support to the Turkish national linux distribution at [TUBITAK](#) (Scientific and Technological Research Council of Turkey, the equivalent of the NSF) to building a scalable infrastructure for storing thousands of gigabytes of user messages at Custommade.

Most of my projects are still used to this day, like my [Google Summer of Code](#) project where I worked on devising and implementing a new animations API for the KDE desktop environment for Linux, which is now [part of the main codebase](#) and has been part of hundreds of thousands of installations. Furthermore, I am a strong believer in open source and I have developed a large portfolio of open source projects over the last few years.

I encourage you to take a look at my [resume](#) (which is attached to the application) to get an idea of the range of my experience and projects.

## Conclusion

In conclusion, I am very interested in pursuing a research-heavy M.S. at WPI.

I hope that I was able to convince you that I have the adequate skills and experience to pursue the many opportunities provided by WPI and that I would be a good fit for the program. Please do not hesitate to contact me for further information.

Mali Akmanalp ([mali@akmanalp.com](mailto:mali@akmanalp.com))