Inspiration: Computing drives scientific discovery and commercial innovation in almost every enterprise from biology to aerospace transportation. More computation power will help researchers to explore and find answers to some of the world's biggest problems in science, engineering, and business. **Ultimately, I want to contribute to accelerating scientific innovation by building high-performance and energy-efficient computer systems (Hardware Software Co-design) for general-purpose applications or specific applications like AI and Biology.** I believe pursuing a Ph.D. program will help me to realize my goal. As initial steps, I have built a **state-of-the-art scheduling method with ML** by inferring SSD performance per IO, exploring multi-modal data on news framing, and developing a web-based system to perform computational framing analysis. My work has been published at top conferences (OSDI 2020 [1], EMNLP 2021 [2], EMNLP 2021 [3]).

Background: I studied Physics - Energy Engineering when I was in college. It is about how to generate electricity from renewable resources such as solar, wind, water, and geothermal. After graduation, I decided to pursue a career in Computer Science because the only assignment that I truly enjoyed in four years of university is programming. Even though my knowledge of CS was scarce, it didn't stop me there. **I built my CS foundation by learning Machine Learning, Computer Architecture, Operating Systems, and others from online resources.** While doing it, I was looking for a research opportunity in CS. I'm grateful my professor connected me to international faculty in the US.

Systems Research: I got an opportunity to do research with **Prof. Haryadi Gunawi from the University of Chicago** to investigate the latency predictability of unpredictable flash storage. Researchers have been devoted to inventing "white-box" and "gray-box" solutions but can't be simply deployed because it depends on the vendors' willingness to adjust the device interface. There are also "black-box" approaches but must pay the cost of extra I/Os. Our group built a new technique: let the device be the device (black box) and do not remodel the file systems or applications, but learn the device pattern. Our approach (LinnOS) has the ability of learning and inferring per I/O speed with high accuracy and minimal overhead using a lightweight neural network.

I played a key role in the LinnOS project. I created a neural network model to predict the incoming I/O speed for the SSD applications. It is challenging because the inference must be accurate and the model shouldn't revoke I/Os that can be served fast. On the other hand, I also need to ensure that our model performs quickly and efficiently. Fast inference depends on input preprocessing, the depth of the layers, neuron complexity, and feature representation. While using deeper models with more features can improve accuracy, it will hurt inference latency and would be too expensive for usage in the I/O layer.

Through numerous design iterations, I made a 3-layer light neural network and used only crucial input features. I found out that features generally considered "crucial" like block offsets, read/write flags, or the long history of writes do not play a significant role. In the end, the most important input features are the latencies of a few recently completed I/Os and the number of pending I/Os when those I/Os and the current, to-be-inferred I/O arrive. Overall, this model beats the latest speculative techniques applied by the top tech companies with a minor inference cost of single-digit microseconds. LinnOS reduces average I/O latencies by 9.6-79.6% with 87-97% inference accuracy and 4-6µs inference overhead for each I/O. Our work became a state-of-the-art method and was published on **OSDI 2020.** After this project, I decided to explore CS research in the AI domain.

NLP Research: I continued my research journey with **Prof. Derry Wijaya from Boston University.** I participated in building a Web-based system, called OpenFraming, for analyzing and classifying news frames in text documents. The main goal was to make it accessible to researchers with or without computational backgrounds from a diverse group of disciplines. I greatly enjoyed building this system demonstration. For future projects, I plan to create online demonstrations for selected research outcomes so everyone can play around with them. This demo was published on **EMNLP 2021.**

Next, I worked on a multimodal project by utilizing BERT and ResNet to predict the news frames from a Gun Violence dataset. Our work is the first to conduct computational multimodal framing analysis. First, I trained and tested the ResNet-50 model utilizing only images from our dataset to predict the frames. The model's prediction was initially unfavorable since it only obtained 42.8% accuracy. Then, I extracted information from the images as background knowledge by using Google Visual API and combined the extracted features with fully connected layers from ResNet. The combined model outperformed the ResNet model by having an 87% accuracy on articles with relevant images. This indicates that adding background knowledge represented by Google Visual API tags contains useful information that is discriminative of the article's frame. From this experiment, I figured out that Multimodal Learning is a crucial topic in order to build smarter AI agents. This work was accepted at **EMNLP 2021.** After having some research on different CS topics, I realized I prefer Systems to AI research.

Systems Research: Currently, I'm working with **Prof. Huaicheng Li from Virginia Tech** to explore Memory Disaggregation. For the initial step, I analyzed Google memory traces using DynamoRIO. This task might sound simple but it requires persistence to dig deeper and creativity to observe from multiple perspectives. In order to learn the pattern inside the workloads, I built my own analysis tools such as plotting bandwidth, creating heatmaps of memory usage, providing important ranges of memory addresses, and others. Converting raw text into visualization is fun, we can get valuable insights and apply them to improve the existing systems. I also enjoy documenting these analysis tools on Github. I remember learning documentation that is not beginner friendly is tough therefore I'm trying to make a simple, straightforward, and easy-to-understand documentation so people can utilize them with ease later.

Future Plan: Modern computing devices are constrained by data movement. The major portion of the total energy consumption is utilized to move data between the CPU and memory. Consequently, data movement bottlenecks have a significant impact on workloads like ML, computational biology, and others because of low cache utilization and large datasets usage. I believe adopting a new paradigm of **near-memory processing is the key to improving the performance and efficiency of future computer systems.** While I'm interested in exploring near-memory processing, memory disaggregation, and MLSys interaction, I'm still open to other possibilities because there are many exciting research problems in CS.

At UW: I strongly believe that doing research with UW faculty can help me to realize my goal. I'm especially interested in advancing storage systems under the guidance of **Prof. Ming Liu, Prof. Andrea Arpaci-Dusseau, or Prof. Remzi Arpaci-Dusseau.** I would also be intrigued by the opportunity to work with **Prof. Matt Sinclair, Prof. Michael Swift, and Prof. Joshua San Miguel.** I'm confident that with my experience and determination, I can actively contribute to your research group, and hope to have the opportunity to grow together within it.

References:

[1] Mingzhe Hao, Levent Toksoz, Nanqinqin Li, **Edward Edberg Halim**, Henry Hoffmann and Haryadi S. Gunawi. ***LinnOS: Predictability on Unpredictable Flash Storage with a Light Neural Network.*** *In the 14th USENIX Symposium on Operating Systems Design and Implementation (OSDI), 2020.*

[2] Isidora Tourni, Lei Guo, Taufiq Daryanto, Fabian Zhafransyah, **Edward Edberg Halim**, Mona Jalal, Boqi Chen, Sha Lai, Hengchang Hu, Margrit Betke, Prakash Ishwar and Derry Tanti Wijaya. ***Detecting Frames in News Headlines and Lead Images in U.S. Gun Violence Coverage.*** *In the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021.*

[3] Vibhu Bhatia, Vidya Prasad Akavoor, Sejin Paik, Lei Guo, Mona Jalal, Alyssa Smith, David Assefa Tofu, **Edward Edberg Halim**, Yimeng Sun, Margrit Betke, Prakash Ishwar and Derry Tanti Wijaya. ***OpenFraming: Open-sourced Tool for Computational Framing Analysis of Multilingual Data.*** *In the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2021.*