Adrian Cheung

(I chose the topic of data science, but since it was rather broad, I focused mainly on machine learning because it captivated me while I was researching!)

Up until a few decades ago, science has always been developed through experimentation and theory. Then, computers came into the picture, allowing for complex models and approximate solutions to previously unsolvable problems. Now, in the rapidly emerging field of data science, computers can detect patterns and gain knowledge from just the statistics we feed them. One of the major applications of data science is machine learning, in which computer systems can improve by themselves over time with data without being given explicit instructions. Machine learning is what powers autonomous vehicles, image recognition, virtual assistants, search results, and even ad targeting. In the Navy, this tech can make complex judgments like detecting enemy forces from a video feed or finding the safest route for a ship, thousands of times faster than a human. But with all this talk about artificial intelligence and machine learning, how does it even work? What makes a computer "smart"?

When a baby says its first word, it is imitating others, having heard the sound before. As it is taught different words while being shown different objects, it begins to associate them and learn the meaning of each word. The more word-object pairs there are, the more the baby learns. Machine learning models, particularly neural networks, behave similarly. In a neural network (the baby), input data (the objects' images) and output data (the corresponding words) are fed into a model, which can be thought of as a very complex mathematical function. This function is first randomly made and generates a hypothetical output from the inputs it receives. Comparing this generated output with the real output data, the error is calculated, telling the model how badly it performed. The model then makes adjustments and generates new outputs. If the error decreases, signaling it's doing better, then it keeps going in that direction; if the error increases, then it tries different adjustments. Almost like a game of hot and cold, repeating this process many times will eventually yield a function that has a low error, meaning it fits the data well. If this sounds familiar, the bare-bones neural network is essentially a regression in statistics, except working with complex data instead of just a couple of points. The complexity of neural networks is also why lots of data and computing power are needed, which was barely feasible just a decade ago.

The heavy reliance on large quantities of data in machine learning calls for lots of data collection, preprocessing, and exploratory analysis beforehand. Data scientists spend most of their time handling such monstrous datasets so that models can analyze more efficiently and effectively. For example, a researcher may be looking into image classification of naval mines. Suppose they notice that there are significantly more images without mines than those with mines, since it may be difficult to get data for the latter. To prevent the model from developing a bias towards the absence of mines, the researcher could add to the existing set of images with mines by augmenting it—in other words, duplicating the images but making different adjustments such as crop, rotation, distortion, brightness, contrast, saturation, and filters. As an added bonus, the model has more data to train on, increasing the accuracy. Of course, data augmentation is just one of the many things data scientists do in their analyses. These vast, rapidly growing methods and applications are why I'm particularly interested in data science and machine learning. Even within the Navy, there are many complex tasks that can be done in this field. Radar signals can be automatically monitored for unusual changes. Scheduling algorithms can be used to optimize when to send data between ships. Swarm tactics can be automatically created to aid military operations. Seafloor properties can be determined and 3D depth maps can be generated using deep neural networks onboard autonomous underwater vehicles. The possibilities are endless!

However, the wide area machine learning encompasses isn't the only reason I am inspired by it and have enjoyed researching it for the past few months. To me, data science and machine learning are like really intricate magic, mindblowing on the outside but also mindblowingly complex behind the scenes. When I delved deeper into topics like backpropagation, convolutional neural networks, recurrent neural networks, generative adversarial networks, unsupervised learning, and reinforcement learning, I was amazed by the inner workings, the mathematical foundation, behind the different architectures I could look into. Especially after Mr. Koe described the importance of data in AI, I couldn't help but try it out myself: I found a ship detection dataset on Kaggle and trained a multiclass ConvNet on it using TensorFlow/Keras in Python. Indeed, data cleaning turned out to be a huge part of the project, and I learned a lot about data science and machine learning along the way. His story about encountering data science by being open to different opportunities also particularly spoke to me, because it was this same openness that helped me discover this program and data science in the first place. Thanks to Mr. Koe, who inspired me to do my small project on ship classification, I found that data science was a great fit (no pun intended) for me: my strong math background, combined with my passion for programming and robotics, will definitely guide me as I continue to explore the subject. Since this has been my favorite topic in computer science so far, I hope to become a researcher in data science and machine learning.

It's almost impossible to imagine what the state of data science will become in the next 15 to 20 years—who knows when the next Deep Blue or AlexNet will come? With several major breakthroughs every year and hundreds of papers per day, this field is evolving crazy fast. I hope to see an increased dependence on fully autonomous systems, since most machine learning applications we have are still aided by humans. This would allow for cars and traffic lights to synchronize with almost no intervention, and replace the need for a sailor's eyes with sensors. I also expect our models to be more integrated, as they currently often start from no information before training and aren't directly connected with other models. This would be a big step towards mimicking human behavior by retaining information in one big model, increasing the efficiency of AI and making training much faster. Some of our tasks today will be problems of the past, including image classification and object detection. Instead, there will be a shift in focus on more abstract problems and robotics, where machine learning could be used to master movement with many joints and parts. Autonomous soldiers and military robots could be commonplace, and large tactical decisions may end up in the hands of AI. That being said, unless there is a different technology that can replace the current paradigms of machine learning, it seems that neural networks are here to stay for a while, as many parts of it are still unexplored. There are downsides to the emerging technologies in data science, however. With so much data infiltrating our lives, privacy will be a growing problem. Manufacturing and transportation will mostly be dominated by robots, displacing millions of jobs short-term. AI could also be used maliciously, from abusing deepfakes to intelligent cyberattacks on the military. Nevertheless, I believe the benefits significantly outweigh the risks. The amazing possibilities in the future of data science and machine learning bring confidence in my path forward.