- Can you please summarize your prior/current work related to the current position?

I am currently working at Boston Children's Hospital, Harvard Medical School (Radiology department). I was responsible for three different projects:

1. Effects of mother's maternal nutrition along with family's demographic information (education, income, etc) on children's neurocognitive outcomes (language comprehension, gross motor, general development)
2. HIE outcome prediction based on both infants clinical information (birth weight , hypothermia, Head circumference, Sex, etc) and MRI brain image
3. 3D MRI Image segmentation

- What is your motivation/interest for the current position?

I would say my last experience at Boston Children's Hospital and academic projects from Johns Hopkins University gave me a lot of interest in medical domains, although I still have a lot of domain knowledge that I need to learn. I would love to apply my machine learning and deep learning experience onto the medical domain. Moreover, my PI graduated from Upenn and he thought Upenn would be a great place to learn and contribute.

- What is your experience in programming/coding/scripting? Which programming language(s), development platform(s), operating system(s) you most commonly use?

I have at least 6 years experience in Python programming. I used numpy library, sk-learn, and Pytorch (deep learning). Linux and Mac OS are the most common operating system i uses.

- Describe two data analysis projects you've actively participated in the past?

Other than internship experiences at Boston Childrens' Hospital, most of my data analysis experience related to application products. In 2019, I worked for a short-video platform company called Kwai which was pretty similar to TikTok. Part of my projects was to retrieve 10 million data every day by using Hive SQL and analyze the data via attribution analysis. For example, If the DAU (Daily Active User) on homepage decreased 10% compared to last week, I thought of the reasons for this happening from different aspects of metrics. First, I checked whether the data was accurate or not. If the data was accurate, then I split possible reasons into different metrics: New/Old users; IOS/Android platform; Different Version due to AB-Test; Different entrances to homepage interface, etc. When I located the problem, I made further splits into product side, technology side and operation side. I set up a group chat for corresponding co-workers and checked what adjustments have been made this week to find the reasons. I also used the topic modeling technique to explore group chat information and check if anyone sent fake news. I used statistical methods to determine whether new features will benefit our application product through AB testing. In the end, the "Follow Page" user penetration rate increased to 70% and user consumption time increased by 3% compared to last quarter.

- Do you have a publicly accessible repository of your projects/code (e.g. GitHub account)? If not, could you send example code?
Yes, you can find them here: https://github.com/dwu12/Machine-Learning-Project

- Briefly describe your knowledge and experience in statistical data analysis, data visualization techniques and machine learning.
Almost every project / work experience needs data analysis, data visualization and machine learning skills. When I get clinical (tabular) data, I will first check whether there are some missing values. Then I can explore the data with descriptive statistics: mean, median, max, min, etc (for example, dataframe.describe() in python). Then I can check the datatype of the target / response / dependent variable. If it is numerical, I can use a pair plot or heatmap to plot the correlation (this could be pearson or spearman correlation). If the target value is categorical, then a counterplot could be used to see whether the data is balanced or not. I can also use hypothesis tests to see whether two dataset is truly different. Those are all related to statistical data analysis. For data visualization, either python and tableau could be used, where tableau is more convenient and python also has pretty good library (matlab plot library and seaborn library.)

For machine learning, I knew most of the commonly used machine learning algorithms and how to robust them with cross-validation and grid-search techniques. Moreover, I have experience with deep learning for both computer vision and natural language processing. Currently I am really interested in contrastive learning, and I hope I can use them in later projects. Since supervised learning is quite expensive, especially in biomedical domains where we need experts to annotate the data, maybe contrastive learning could benefit medical areas.

- In addition to excellent technical skills this position also requires strong organizational skills. Could you briefly describe/comment on your organizational skills?
The way I keep everything efficiently and effectively is to set up a regular meeting in order to make sure everyone is on the same page. Personally, I feel like meeting in person is more efficient than chatting by email. Also, I love to write a doc for the details of each version I made including the code so that I can easily keep track of the progress. When there are multiple projects on the same time, I will split them into small projects in a specific timeline with priority that can help meet my goals.

- Do you have experience working with medical imaging? Please describe briefly, if you do.
Yes, I have worked with MRI medical imaging for both image segmentation and image classification. For image segmentation I compared 3D Unet and 3D Unet transformer, however, I found that under the same conditions, when the data is not quite large, the transformer model will not outperform traditional CNN.