大学生创新创业训练项目是培养学生具有创新能力的重要环节,目前全国很多学校把大学生参加科研训练作为提高大学生科研素质的重要途径。作为一名东大学子，抱着来为求知，去做栋梁的信念，我在大二参与了本次srtp项目——面向医学文本的多标签分类系统。在为期一年的项目过程中，我从懵懂的计算机小白，成长为能够与研究生学长学姐一同商讨自然语言处理知识的组员，向着科研之路迈进了一大步。

通过此次科研，我学习到了人工智能一大领域——自然语言处理的许多知识。自然语言处理 (Natural Language Processing, NLP) 是计算机科学，人工智能，语言学关注计算机和人类（自然）语言之间的相互作用的领域，语音识别、智能问答、情感分析等领域都能见到自然语言处理的身影。自然语言是人类智慧的结晶，而自然语言处理是人工智能中最为困难的问题之一，而对自然语言处理的研究也是充满魅力和挑战的。我们的科研训练项目围绕自然语言处理在医学文本上的应用，主要目标是对一片文献中的大量语句进行分类添加标签，再加以抽取和提炼，将一篇冗长的科研论文浓缩成一份简短的概要，极大地提高了研究人员阅读文献的效率，帮助他们更好的进行科研工作。在项目进行的过程中，我们阅读了许多论文，学习了关于医学界通用模型的知识，多标签分类的知识、深度学习的一些常用框架、最终挑选lstm、bert两种神经网络模型作为我们分类的方法。

通过此次科研训练，我体验了科研从理论到实践的过程，并深刻认识到“实践是检验真理的唯一标准”的重要意义。一个好的科研项目，不仅仅需要扎实的基本功，灵活的思维，更需要将想法与实践相结合。虽然我们的核心思想是对医学文本进行分类，但是为了将理论变为现实，我们必须进行一套完整的自然语言处理流程。在项目进行过程中，我们从最初的制作数据集，到文本预处理，再到搭建神经网络模型，使用数据集训练模型，到最终的结果预测，每一步都进行了仔细的研究。与此同时，顺手的研究工具也不可或缺。我们学习了tensorflow和keras等人工智能研究常用工具和框架的使用方法。将理论转化为实践是一个鼓舞人心的过程，当看到一句未知的文本被正确分类，搭建的模型成功实现了自己的想法，那一刻身心的疲倦仿佛一扫而空，我想，科研的魅力就在于此。

有言道：“台上一分钟，台下十年功”。成功的背后是日积月累的辛勤努力。通往成功的道路上也会遇到许多问题。在我们的科研训练过程中，由于初识自然语言处理，标注训练集比较枯燥也很容易出错，有些同学将医学文献编号 pubmed 拼写成了 pembud ；将 disease 拼写成了 disaese ；或者在标签类别中有一些空格等等。还有一些其他问题，比如一开始不知道如何进行文本处理、训练参数设置错误等等。这些问题，都导致了我们的工作出现了困难，但随着我们的不断进步，一段时间后我们也逐一修复了这些 bug，解决了这些问题。

通过此次科研训练，我获益匪浅。它不仅使我对人工智能的一大分支——自然语言处理有了更深入的认识，同时锻炼了我的实践能力，并且结识了三位志同道合的朋友和一位优秀的指导老师，培养了我团队合作的能力。这次科研训练，不仅激发了我对人工智能的浓厚兴趣，更加坚定了我投身科研的热情。