moeCTF2023 AI安全入门指北

Author: Kyriota

AI安全主要指对抗攻击(adversarial attack)类别的模型安全问题，在CTF中，AI相关的题目大多存在于Misc类别下作为一种考察选手对机器学习工具与相应数学概念理解的题目类型。由于参加CTF赛事的选手多为安全方向专业/职业，此类题目通常得分率与关注度较低，但也正因如此，拿下一些关键的AI题往往能扭转乾坤。接下来就让我们用一些`Q&A`简明扼要谈谈AI安全。

Q&A

Q：学AI对电脑的配置要求高吗？

> 机器学习中的计算是通用计算，没有独显的电脑也可以使用CPU算，新手入门不会遇到过大的数据量和模型，就算新手有一块优秀的GPU，他们通常也处理不好GPU配置与调用，所以结论是对入门的配置要求低，后期性能不够用可以按需更换电脑。
>
> 注：本次moeCTF中的AI题中GPU的参与是非必要的

Q：环境配置？

> 目前的主流机器学习库有：`TensorFlow`，`PyTorch`和`Keras`。我推荐`PyTorch`，`PyTorch`几乎是现在学术界的标配，且拥有相比其他两个库更大的用户量，这意味着更多资料
>
> 新手可以先安装`Python`，`VSCode`等，然后安装不带GPU支持的PyTorch（即便你有GPU也不要折腾GPU支持）就可以开始实践了

Q：我需要有怎样的基础？

> 机器学习涉及到：Python，高等数学，线性代数，炼金术，英语阅读理解（选修）等，对于数学板块，与密码学Crypto主打数论不同，AI中的数学大多是“非常具体的数学”，很多与AI相关的知识都可以做到充分的、易于理解的可视化，如果你对三维空间、立体几何拥有十足的洞察力，那不妨来挑战一下维度更高的机器学习的世界。
>
> 注：本次moeCTF中的AI题不会涉及到对23级新生来说过于困难的数学问题，也不会涉及到炼金术

Q：学习AI的发展前景？

> 一个优秀的AI基础在数模竞赛与科研中有着一锤定音的关键效果。AI的分支方向非常多，在你搜索到的每一个AI细分方向下都养着一大窝的科研狗，而由于AI中炼金要素的存在，在AI方向水论文，发表学术成果是相对容易的。

Q：我准备好了，怎么开始学习AI安全？

> AI安全涉及到的对抗攻击可以说是AI训练过程的一个反向，因此，如果你希望掌握机器学习中的对抗攻击，你需要先掌握在这之前的一切，包括但不限于：反向传播，偏导数，Python进阶，常用的机器学习模型类型等。
>
> 但限于本次CTF的难度控制与出题人的勤快程度，本次moeCTF首批AI题并无对抗攻击类题目

Q：推荐的学习资料？

> 吴恩达机器学习：经典古董级教程，适合用于打下坚实的数学与概念基础。
>
> YouTube 李宏毅：台大教授，ACG爱好者，我推的教授，适合用于在基础好的前提下快速入门新知识或了解AI圈子前沿资讯解读。
>
> PyTorch官方：PyTorch Tutorial，PyTorch Document

Q：本次moeCTF的AI题难度与出题方向？

> moeCTF的目的是入门和兴趣，而非使人抑郁，故为了保证题目难度上升梯度平缓，以及照顾到23级新生的知识体系，本次AI题主要是关于机器学习基本知识与Misc的杂糅，而不是传统意义上的AI对抗题。尽管如此，个别题目的预想难度仍然不低。如果后续AI分区做题反响可观，将会根据比赛剩余时间考虑比赛中途补充对抗攻击的题目。

Q：我Flag呢？

> Flag = `moectf{A_B_C_D}`，官方是建议的不带GPU支持的PyTorch安装指令是：
>
> `pip3 ___A___ ___B___ ___C___ ___D___`
>
> 将`A`、`B`、`C`、`D`在指令中对应的单词代换到Flag中即可（完形填空），保留下划线作为分隔
> Example Flag：`moectf{this_is_an_example}`