1. **总结**

在本文中，我们详细介绍了基于AI的PDF恶意文档分类器的设计与实现。实验数据表明，在十万级文档数据集中，我们能达到大于99%的准确率和小于0.01%的误报率。且在实际运行时，CPU和内存的时空效能比（Time & Space Performance）比旧有基于规则的模型，有显著提升。

本文除了使用大量数据研究人工智能化应用安全，并且把人工智能本身的安全也放在了同等重要的位置上。我们通过大量的实验，（1）模拟了攻击者通过对恶意样本的增删改（如变更特征的值），以混淆分类器，达到逃逸的目的；（2）分类器经自身修正，通过重新训练模型，去除已被攻击者所利用的特征，以维持模型的健壮性。

基于AI的文档分类器是社会工程学、病毒分析等领域的重要研究课题。在未来，我们还会尝试解决以下研究问题：

* 基于深度学习的恶意PDF文档检测
* 动静态分析引擎的调优
* 对于Microsoft Office等其他文件格式的支持（如docx，pptx等）

Conclusion

In this paper, we introduce the design and implementation of PDF file classifier based on machine learning. Our experiment results reveal that in the dataset at a hundred level, we can achieve a detection accuracy which is larger than 99% and a false positive rate that is less than 0.01%. Besides, compared with the model based on rules, the Time and Space Performance of CPU and memory is improved significantly when the classifier is running in real time.

We use a great amount of data to study the application security based on artificial intelligence and at the same time give equal weight to the security of artificial intelligence itself. Through a lot of experiments, we have done the following works: 1) attackers change the malicious sample (such as modify the value of features) so as to confuse classifier and then can evade it; 2) the classifier conduct self-revision by retraining the model and eliminating the features that have been exploited by attackers in order to maintain the model robustness.

The learning-based classifier is an important research topic in the field of social engineering, malware analysis, etc. In the future, we will try to focus on the following topics:

* Malicious PDF file detection based on deep learning
* Optimization of static and dynamic analysis engine
* More file formats such as Microsoft Office including docx, pptx, etc.