

流智图灵—基于轻量预训练模型与多源知 识融合的威胁感知系统



CONTENT

F l o w w i s e T u r i n g s

随着互联网的快速发展和数据的指数级增长，网络安全已成为全球面临的重大挑战。企业和个人每天都产生大量的数据，这些数据中可能含有敏感信息，而存储着这些敏感信息的系统则会成为黑客攻击的目标。同时，网络威胁日益复杂多变，传统的安全防护措施难以应对新型的网络攻击手段，大量因为攻击而产生的流量数据也使得防守任务变得异常冗杂。在这样的背景下我们的流智图灵应运而生。



Part one

项目背景



Part two

项目概述



Part three

成果展示



Part four

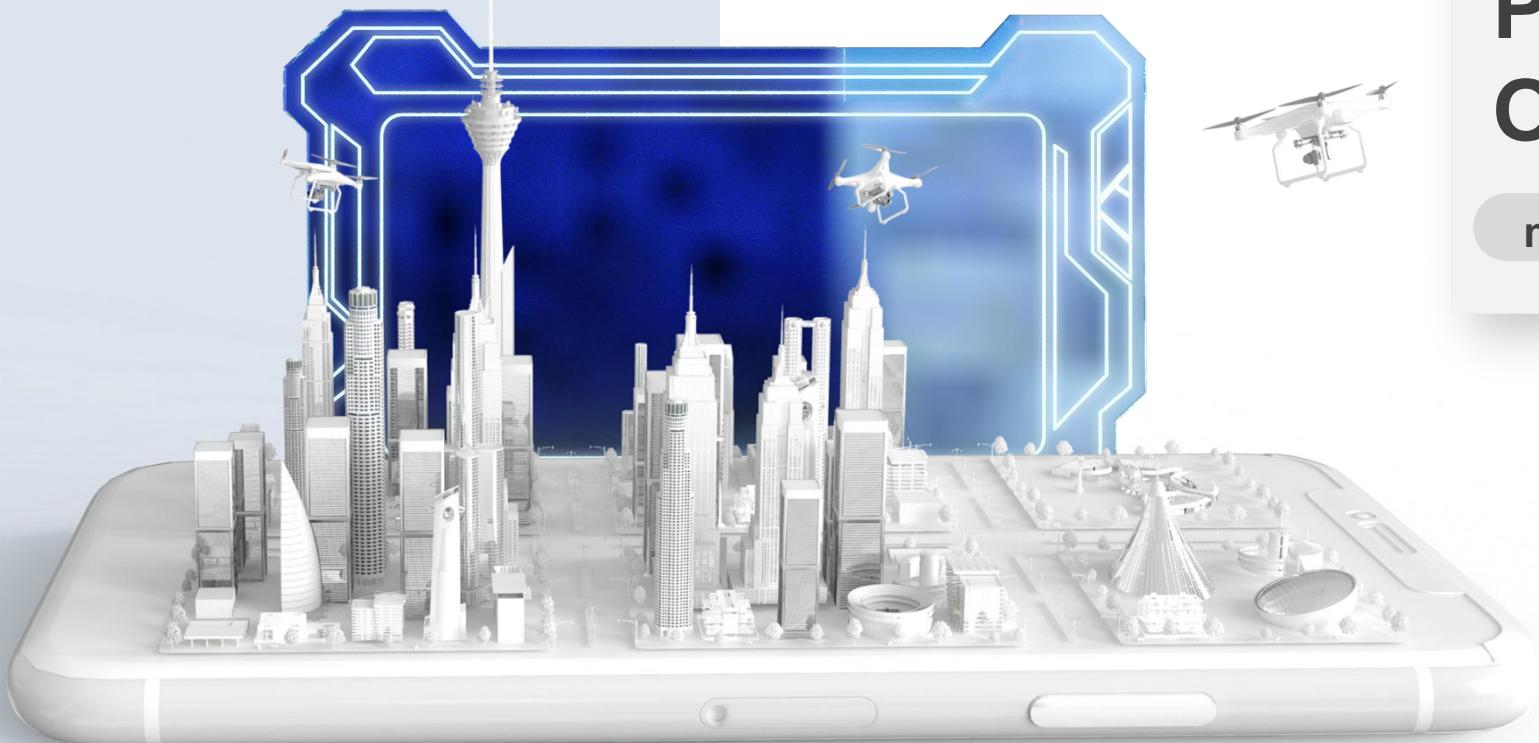
详细技术



Part five

项目总结

项目背景



PART
ONE

more

行业背景 /Industry background

17

2024



针对我国网络安全和信息化发展中遇到的新形势新挑战新问题，以习近平同志为核心的党中央总揽全局、沉着应对、高瞻远瞩、把舵领航，党的二十大站在推进国家安全体系和能力现代化的战略高度，对完善国家网络安全治理体系作出新的部署，纵深推进我国网信事业发展繁荣有序、稳步前行。

2023



“网络安全牵一发而动全身”，“没有网络安全就没有国家安全”……党的十八大以来，习近平总书记深刻把握信息化发展大势，高度关注网络安全挑战，多次在不同场合就网络安全发表一系列重要论述，把我们党对网络安全的认识提升到了新的高度和境界，为树立正确的网络安全观、做好网络安全工作提供了根本遵循和强大动力。我们摘录了习近平总书记有关重要论述，与您一起学习网络安全的重要性。

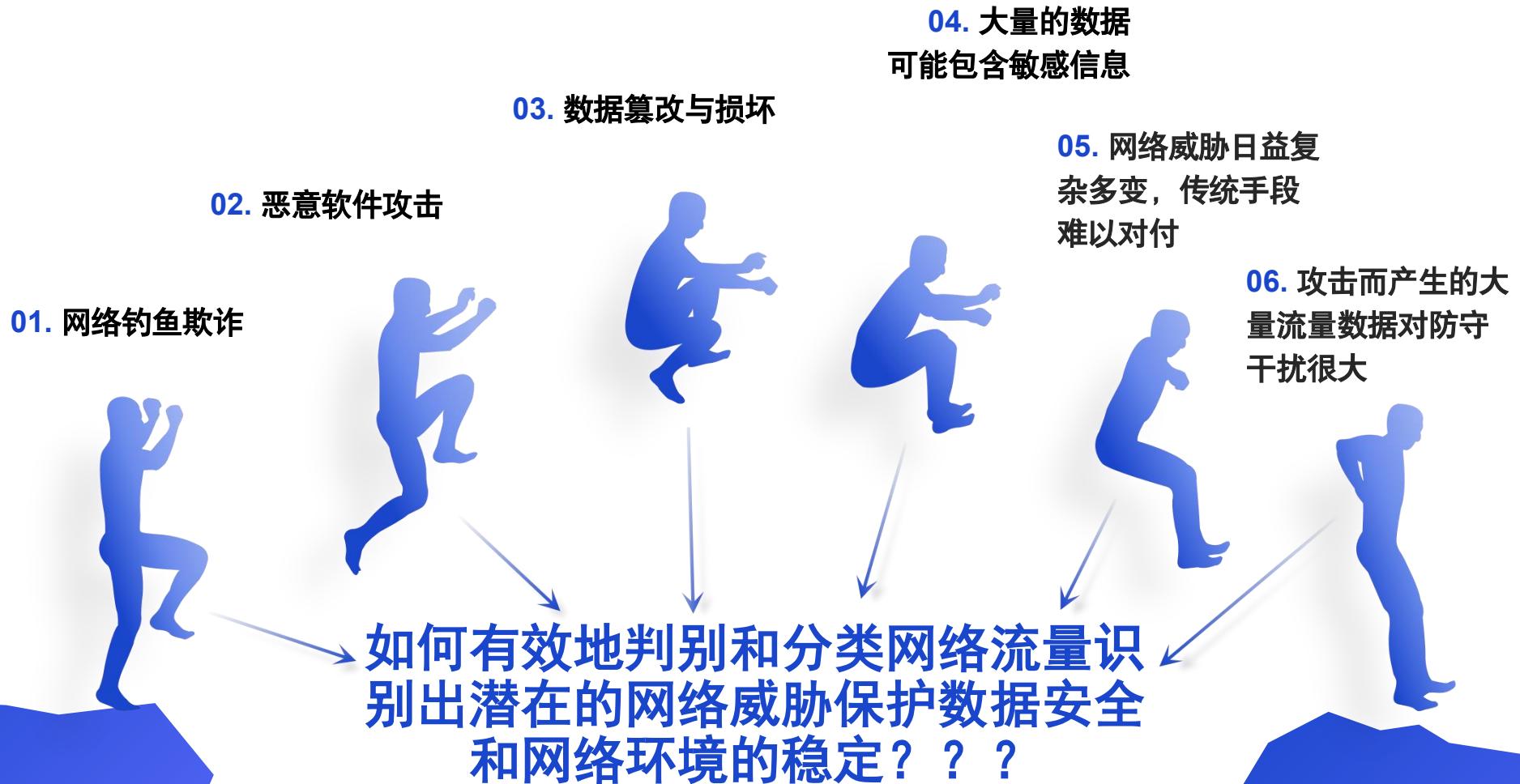
About our FlowwiseTurings

关于我们的“流智图灵”

而在2025年《政府工作报告》中，一个新关键词引发热议——“人工智能+”行动，这是“人工智能+”首次被写入政府工作报告。在今年全国两会中，多位来自网络安全领域的专家都在提案中建议创新发展“AI+安全”，以提高应对网络空间安全风险与不确定性的能力，于是，一个高度先进的威胁感知平台——流智图灵，应运而生。

2025

网络安全现状与挑战



需求分析 /Requirements



先进的人工智能技术



(1) 系统流量判别功能与网络流量分类功能需要结合先进的人工智能技术，如RoBERTa-MiniLM和LSTM，AI模型将精准的进行威胁识别和分类，以帮助用户迅速了解流量的性质和潜在风险；



知识图谱，高速分析处理能力



(2) 智能问答系统需结合知识图谱技术，提供深度、可定制的问答服务，帮助用户准确理解和应对网络安全问题；

(3) 系统需要实现高效的数据处理能力，特别是在处理大量网络流量数据时，能够保持高速的分析和判断能力，确保网络环境的实时监控和防护；



强大的数据保护和隐私保障机制



(4) 必须具备强大的数据保护和隐私保障机制，确保所有监测和处理的网络流量数据安全，防止数据泄露或被非法利用，从而维护用户和企业的信息安全。



目标用户

Target user

在21世纪，随着互联网的快速发展和数据的指数级增长，网络安全已成为全球面临的重大挑战。企业和个人每天都产生大量的数据，这些数据中可能含有敏感信息，而存储着这些敏感信息的系统则会成为黑客攻击的目标。同时，网络威胁日益复杂多变，传统的安全防护措施难以应对新型的网络攻击手段，大量因为攻击而产生的流量数据也使得防守任务变得异常冗杂。在这样的背景下，如何有效地判别和分类网络流量，识别出潜在的网络威胁，保护数据安全和网络环境的稳定，成为了亟需解决的问题。

(1) 网络安全管理员和分析师：负责监控和维护网络安全的专业人员，使用流智图灵可以帮助他们识别和防御网络威胁，提高安全监控的效率和准确性；



@网络安全管理员和分析师

(3) 网络安全研究人员：从事网络安全研究的学者和技术专家，使用流智图灵可以帮助他们进行网络威胁分析和研究，加深对网络安全问题的理解和掌握；



@网络安全研究人员

(2) 企业和组织的IT部门：包括需要维护网络安全和数据保护的企业、政府机构和非营利组织，使用该系统可以强化他们的网络防御能力，保护组织免受网络攻击和数据泄露；



@企业和组织的IT部门

(4) 技术支持和咨询服务提供商：为企业或个人提供网络安全支持和咨询服务的机构，使用该系统可以增强服务能力，为客户提供更有效的安全解决方案。



@技术支持和咨询服务提供商

PART TWO

项目概述

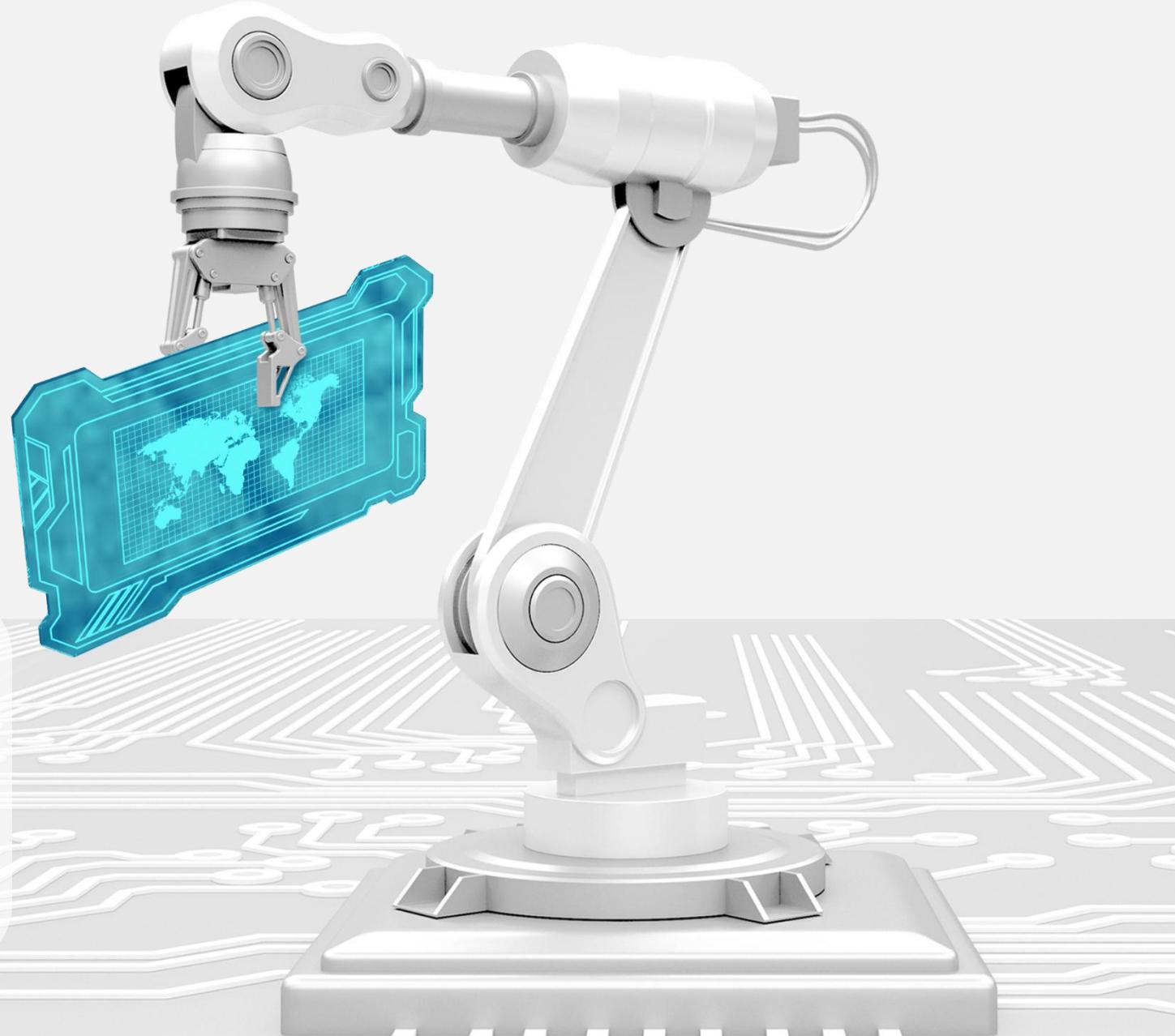
Project Overview



“没有网络安全 就没有国家安全”

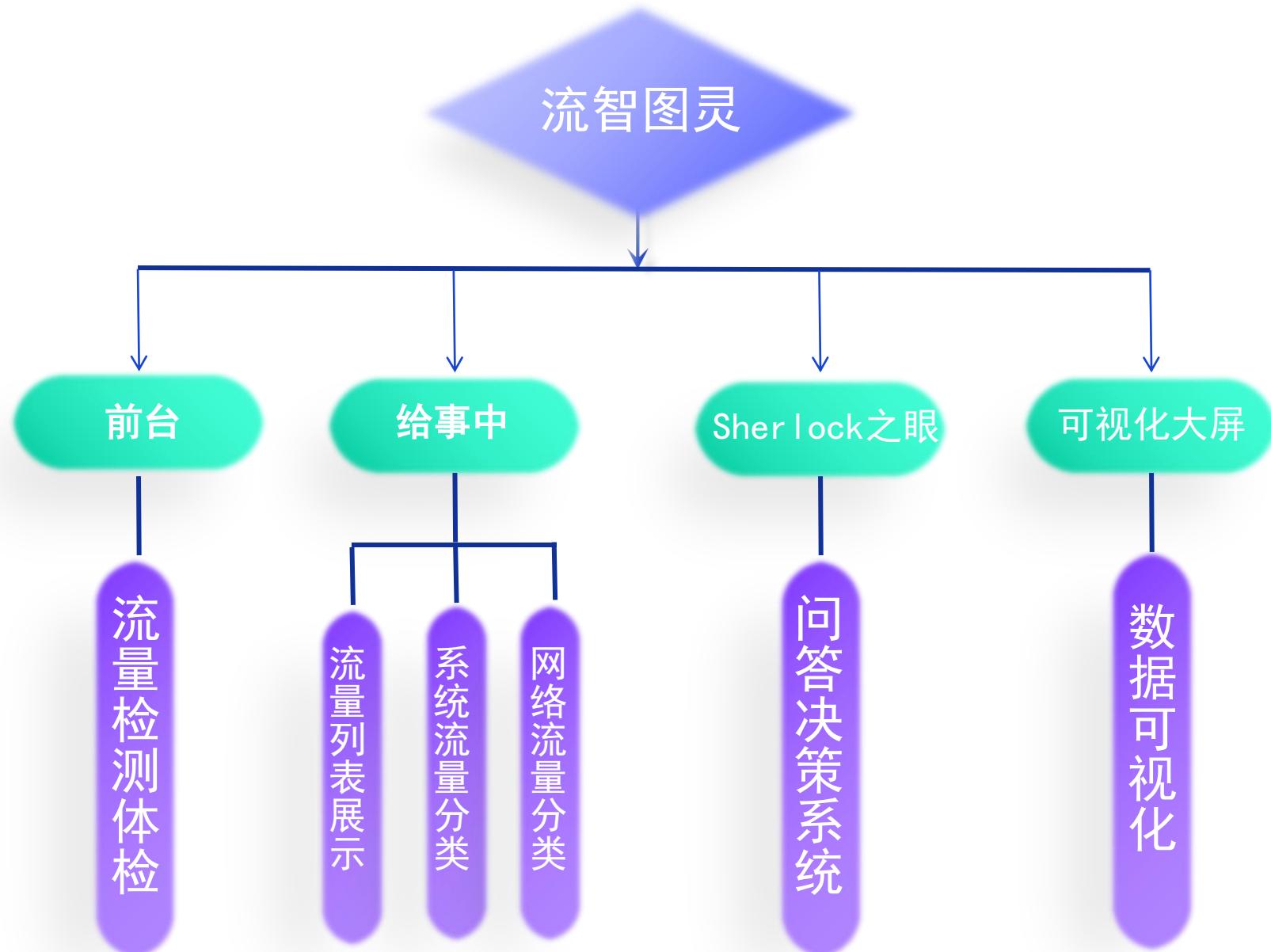


Without cybersecurity
there is no national security.



FlowwiseTuring

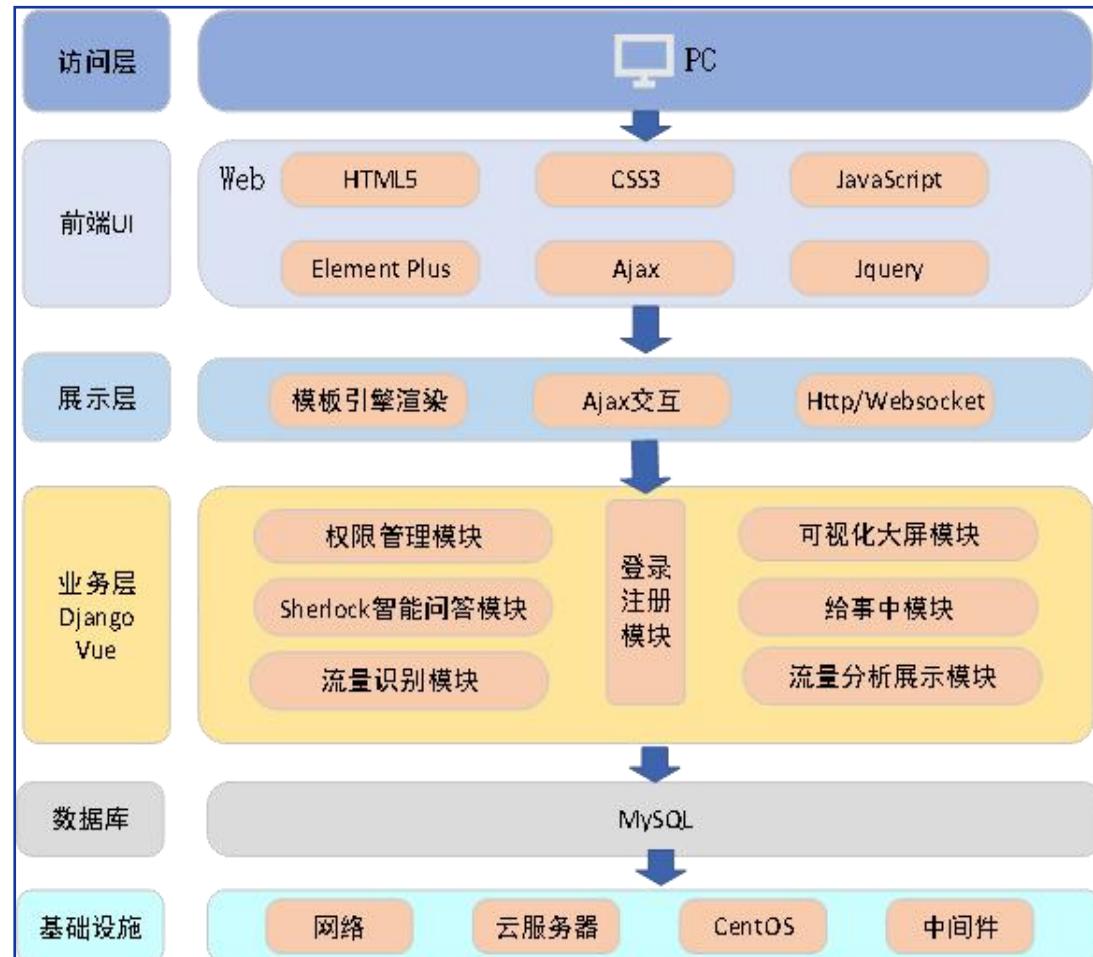
流智图灵





系统设计架构

- Element组件库，人机交互好
- Ajax交互，减轻系统负担，提高性能
- 基于Django和Flask框架，MTV模式，低耦合
- 云数据库存储信息
- 跨站点请求伪造(CSRF)保护，安全性更高



Django

Element组件库

Ajax交互



总体设计

流智图灵设计完成了给事中、Sherlock之眼、数据大屏可视化和用户权限管理功能。依托深度学习模型如RoBERTa-MiniLM和LSTM及知识图谱技术，确保高效准确的威胁识别和网络安全管理。给事中分析判别各类型流量，API对接设备实现自动告警。Sherlock之眼提供智能问答融合知识图谱，辅助决策和应急响应。可视化大屏展示系统运行和流量统计。系统提供登录、权限管理，密码云存储加密，保障安全。用户交互简便。



PART THREE

成果展示

Test report

- 系统流量判别功能测试
- 网络流量分类功能测试
- 知识图谱功能测试





```
  .name": "php_",
  "proc.name": "php-fpm",
  "proc.name": "mysqldadmin",
  "proc.name": "php-fpm", "pid": "72", "evt.dir": "<",
  "proc.name": "mysqldadmin", "pid": "72", "evt.dir": ">", "ev
  8", "proc.name": "mysqldadmin", "l .vpid": "72", "evt.dir": "<", "evt.
  14", "proc.name": "php-fpm", "proc.pid": "89", "evt.dir": "<", "evt.type
  60", "proc.name": "mysqldadmin", "p .c.vpid": "72", "evt.dir": "<", "evt.ti
  98", "proc.name": "php-fpm", "proc.vpid": "89", "evt.dir": "<", "evt.type".
  29", "proc.name": "php-fpm", "proc.vpid": "55", "evt.dir": "<", "evt.type":
  46", "proc.name": "sh", "proc.vpid": "1", "evt.dir": "<", "evt.type": "fork
  54", "proc.name": "nginx", "proc.vpid": "65", "evt.dir": "<", "evt.type": "
  18", "proc.name": "sh", "proc.vpid": "75", "evt.dir": ">", "evt.type": "ope
  63", "proc.name": "nginx", "proc.vpid": "67", "evt.dir": "<", "evt.type": "
  31", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": "<", "evt.type": "
  25", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": ">", "evt.type": "
  81", "proc.name": "ps", "proc.vpid": "73", "evt.dir": "<", "evt.type": "ope
  78", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": "<", "evt.type": "
  62", "proc.name": "ps", "proc.vpid": "73", "evt.dir": "<", "evt.type": "ope
  57", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": ">", "evt.type": "
  83", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": "<", "evt.type": "
  95", "proc.name": "ps", "proc.vpid": "73", "evt.dir": ">", "evt.type": "ope
  68", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": ">", "evt.type": "read",
  "evt.arg0": "<4t>127.0.1:50904->127.0.0.1:9000", "evt.arg1": "O_RDONLY|O_CLOEXEC", "ev
  43", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": "<", "evt.type": "read",
  "evt.arg0": "8", "evt.arg1": ".....", "evt.arg2": "<NA>" "evt.arg0": "<4t>127.0.0.1:50904->127.0.0.1:9000", "evt.arg1": "O_TRUNC|O_CREAT|O_WRONLY",
  65", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": ">", "evt.type": "read",
  "evt.arg0": "<4t>127.0.0.1:50904->127.0.0.1:9000", "evt.arg1": "..QUERY_STRING..REQUEST_MET
  04", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": "<", "evt.type": "read",
  "evt.arg0": "656", "evt.arg1": "..QUERY_STRING..REQUEST_MET
  20", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": ">", "evt.type": "read",
  "evt.arg0": "<4t>127.0.0.1:50904->127.0.0.1:9000", "evt.arg1": "O_RDONLY|O_CLOEXEC", "ev
  12", "proc.name": "ps", "proc.vpid": "73", "evt.dir": ">", "evt.type": "read",
  "evt.arg0": "<f>/proc/1/cmdline", "evt.arg1": "2003", "evt.arg2": "
  07", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": "<", "evt.type": "read",
  "evt.arg0": "8", "evt.arg1": ".....", "evt.arg2": "<NA>" "evt.arg0": "<4t>127.0.0.1:50904->127.0.0.1:9000", "evt.arg1": "O_RDONLY|O_CLOEXEC", "ev
  03", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": ">", "evt.type": "read",
  "evt.arg0": "8", "evt.arg1": ".....Y..", "evt.arg2": "<NA>" "evt.arg0": "<4t>127.0.0.1:50904->127.0.0.1:9000", "evt.arg1": "O_RDONLY|O_CLOEXEC", "ev
  13", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": "<", "evt.type": "read",
  "evt.arg0": "8", "evt.arg1": ".....Y..", "evt.arg2": "<NA>" "evt.arg0": "<4t>127.0.0.1:50904->127.0.0.1:9000", "evt.arg1": "O_RDONLY|O_CLOEXEC", "ev
  15", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": ">", "evt.type": "read",
  "evt.arg0": "8", "evt.arg1": ".....Y..", "evt.arg2": "<NA>" "evt.arg0": "<4t>127.0.0.1:50904->127.0.0.1:9000", "evt.arg1": "O_RDONLY|O_CLOEXEC", "ev
  5", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": "<", "evt.type": "read",
  "evt.arg0": "89", "evt.arg1": "id=1%27+%7C%7C+$28select+fla
  ", "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": ">", "evt.type": "read",
  "evt.arg0": "<4t>127.0.0.1:50904->127.0.0.1:9000", "evt.arg1": "O_RDONLY|O_CLOEXEC", "ev
  "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": "<", "evt.type": "read",
  "evt.arg0": "", "evt.arg1": ".....", "evt.arg2": "<NA>" "evt.arg0": "<4t>127.0.0.1:50904->127.0.0.1:9000", "evt.arg1": "O_RDONLY|O_CLOEXEC", "ev
  "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": ">", "evt.type": "read",
  "evt.arg0": "<4t>127.0.0.1:50904->127.0.0.1:9000", "evt.arg1": "O_RDONLY|O_CLOEXEC", "ev
  "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": "<", "evt.type": "open",
  "evt.arg0": "<f>/var/www/html/config.php", "evt.arg1": "O_RDONLY|O_CLOEXEC", "ev
  "proc.name": "php-fpm", "proc.vpid": "70", "evt.dir": ">", "evt.type": "write",
  "evt.arg0": "97", "evt.arg1": "NOTICE: PHP message: PHP Nc
  .vpid": "70", "evt.dir": ">", "evt.type": "write", "evt.arg0": "<4t>127.0.0.1:50904->127.0.0.1:9000", "evt.a
  ".: "73", "evt.dir": ">", "evt.type": "open", "evt.arg0": "/proc/55/cmdline" "evt.arg1": "O_RDONLY", "evt.a
  ".: "73", "evt.dir": "<", "evt.type": "open", "evt.arg0": "<f>/proc/55/cmdli
  .vpid": "70", "evt.dir": "<", "evt.type": "write", "evt.arg0": "0", "evt.arg1": "O_WRONLY", "evt.a
  .pid": "67", "evt.dir": ">", "evt.type": "write", "evt.arg0": "<4t>/var/log/n
  ".: "73", "evt.dir": ">", "evt.type": "read", "evt.arg0": "<f>/proc/55/cmdli
  id": "67", "evt.dir": "<", "evt.type": "write", "evt.arg0": "118", "evt.a
  .vpid": "71", "evt.dir": "<", "evt.type": "accept", "evt.arg0": "<4t>127.0.
  "73", "evt.dir": "<", "evt.type": "read", "evt.arg0": "0", "evt.arg1": "O_RDONLY", "evt.a
  ".: "73", "evt.dir": ">", "evt.type": "read", "evt.arg0": "<4t>127.0.0.1:50904->127.0.0.1:9000", "evt.arg1": "O_RDONLY|O_CLOEXEC", "ev
  "3", "evt.dir": "<", "evt.type": "read", "evt.arg0": "<4t>127.0.0.1:50904->127.0.0.1:9000", "evt.arg1": "O_RDONLY|O_CLOEXEC", "ev
  "89", "evt.dir": ">", "evt.type": "read", "evt.arg0": "<4t>127.0.0.1:50904->127.0.0.1:9000", "evt.arg1": "O_RDONLY|O_CLOEXEC", "ev
  "vt.dir": ">", "evt.type": "read", "evt.arg0": "8", "evt.arg1": ".....Y..", "evt.arg2": "<NA>" "evt.arg0": "<4t>127.0.0.1:50904->127.0.0.1:9000", "evt.arg1": "O_RDONLY|O_CLOEXEC", "ev
  ".
```



测试技术指标

F1 score是精确率（Precision）和召回率（Recall）的调和平均，是一系列用于文本分类和多分类任务结果质量的技术指标，广泛用于衡量分类模型的性能。

宏平均

首先对每个类别独立计算F1 Score，然后计算所有类别F1 Score的算术平均。因此所有类别被赋予同等重要性，不用考虑每个类别的样本数量。

微平均

先计算所有类别的累积真正例（TP）、假正例（FP）和假负例（FN），然后用这些累计值计算F1 Score。

加权平均

该步骤与宏平均一致，但在计算F1 Score时，会根据每个类别的样本数量来加权。因此样本量越大的类别对最终F1 Score分数影响更大。



技术指标

$$F1\text{score} = \frac{2PR}{(P+R)}$$

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$



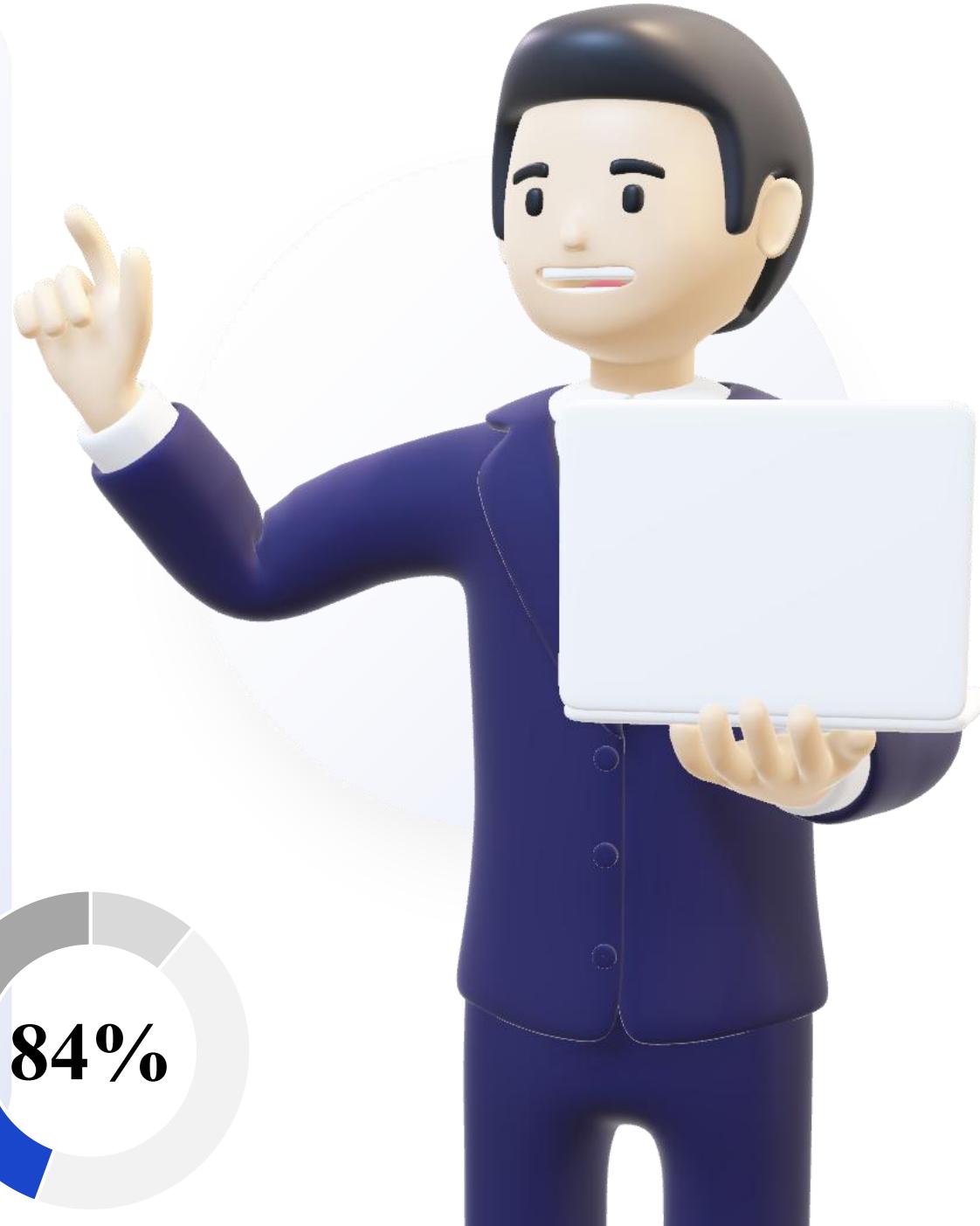
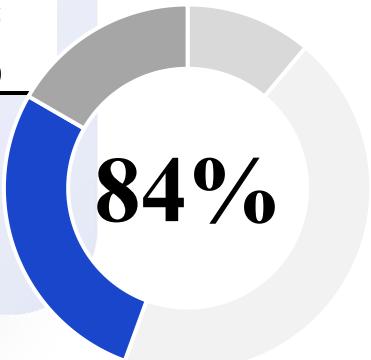
其中，TP表示预测为攻击的日志条目确实为攻击日志条目的数量，FP表示错误地将正常的日志条目预测为攻击日志条目的数量，FN表示错误地将攻击的日志条目预测为正常日志条目的数量。

经过试验，我们的模型再尝试辅以不同的特征工程（`doc2vec`、`fasttext`、`sentence-transformer`）和算法（逻辑回归、随机森林、K最临近、支持向量机、决策树）相结合在测试集（4W多条）

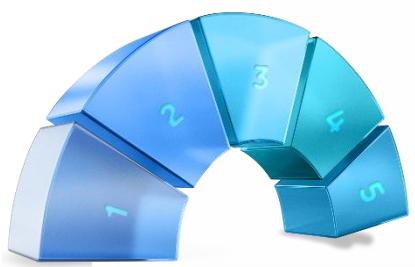
表 4-1 实验结果

| 特征工程+算法 | 精确率(%) | | 召回率(%) | | F1-Score | | F1 Score |
|--------------------------|--------|------|--------|------|----------|------|----------|
| | 1 | 0 | 1 | 0 | 1 | 0 | |
| Sentence-transformer+KNN | 0.64 | 0.94 | 0.97 | 0.48 | 0.80 | 0.78 | 0.79 |
| Fasttext+RandomForest | 0.64 | 0.95 | 0.97 | 0.48 | 0.78 | 0.64 | 0.75 |
| Sentence-transformer+SVC | 0.64 | 0.97 | 0.98 | 0.46 | 0.77 | 0.62 | 0.72 |
| Doc2Vec+ Logistic | 0.63 | 0.85 | 0.91 | 0.48 | 0.75 | 0.62 | 0.70 |
| Fasttext+DecisionTree | 0.67 | 0.96 | 0.98 | 0.53 | 0.79 | 0.68 | 0.75 |
| Fasttext+KNN | 0.63 | 0.87 | 0.92 | 0.48 | 0.75 | 0.62 | 0.70 |

最后我们经过实验得到的数据进行对比，选定了Sentence-transformer+KNN，经过参数搜索，最终F1 Score来到0.84。



系统流量判别



输入一条系统攻击流量数据如下

```
{  
    "@timestamp": "default",  
    "@version": "1",  
    "proc.env": "default",  
    "team_id": "default",  
    "fields": {  
        "fields_under_root": "True",  
        "filetype": "sysdig"  
    },  
    "evt.datetime": "2023-11-12 13:30:57.058843677",  
    "proc.name": "php-fpm",  
    "proc.vpid": "67",  
    "evt.dir": ">",  
    "evt.type": "read",  
    "evt.arg0": "<4t>127.0.0.1:45942->127.0.0.1:9000",  
    "evt.arg1": "8",  
    "evt.arg2": "<NA>",  
    "evt.arg3": "<NA>",  
    "fd.name": "127.0.0.1:45942->127.0.0.1:9000",  
    "proc.ppid": "51",  
    "proc.exepath": "/usr/local/sbin/php-fpm",  
    "evt.rawres": "<NA>",  
    "fd.lip": "127.0.0.1",  
    "fd.rip": "127.0.0.1",  
    "fd.lport": "45942",  
    "fd.rport": "9000",  
    "container.id": "af7ba95e745b",  
    "container.name": "default"  
}
```



相关情报

| | | | | | |
|----------------|-------------------------|---------------|-----------|------------|--------|
| 文件名(File) | systemtest | 请求方式(Request) | read | 攻击类型(Type) | Attack |
| 异动目录(Abnormal) | /usr/local/sbin/php-fpm | 攻击者(rip) | 127.0.0.1 | | |

攻击行为画像



模型给出的判别为：Attack，即模型判断此条流量为攻击流量，并且在系统中给攻击者行为做出了描述如上图。通过与数据集标记对比得出，本系统的系统流量判别功能效果较好，与数据集的标记基本一致，具有极佳的判别能力。



网络流量分类

中国大学生计算机设计大赛 - jzjds.bjtu.edu.cn

流智图灵
基于深度学习的网络安全平台

上传文件分析

数据包分析

```
POST /Autodiscover/Autodiscover.xml !DOCTYPE xxe
[ ! ELEMENT name ANY >
<! ENTITY xxe SYSTEM "&file : //etc/passwd;">] Autodiscover
<xmlns "http://schemas.microsoft.com/exchange/autodiscover
/outlook/responseschema/2006a">
<Request>
<EMailAddress>aaaaa</EMailAddress>
<AcceptableResponseSchema>xxe</AcceptableResponseSchema>
</Request>
</Autodiscover>
```

分析

大赛主页 | 系统简介 | 用户协议 | 联系我们



- 对网络流量分类所做的实验，我们准备了一条XXE（XML External Entity）的网络攻击流量数据

经过人工评判，网络流量分类功能效果良好，能够有效帮助用户检测出攻击流量类型。

```
POST /Autodiscover/Autodiscover.xml !DOCTYPE xxe
[ ! ELEMENT name ANY >
<! ENTITY xxe SYSTEM "&file : //etc/passwd;">] Autodiscover
<xmlns "http://schemas.microsoft.com/exchange/autodiscover
/outlook/responseschema/2006a">
<Request>
<EMailAddress>aaaaa</EMailAddress>
<AcceptableResponseSchema>xxe</AcceptableResponseSchema>
</Request>
</Autodiscover>
```



PART FOUR

详细技术

core technology



Process One

关键技术

建立了两个模型，其一是用于识别解析系统流量的系统流量判别模型（System Flow Discriminate Model, SFDM），其二是用于分辨网络流量的网络流量分类模型（Network Flow Classificate Model, NFCM）。



Process Two

界面设计

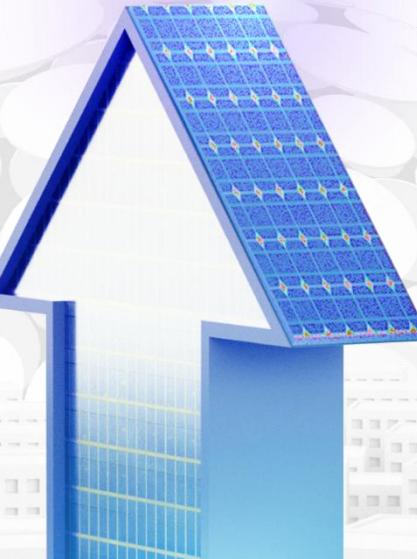
本系统前端页面的实现主要是利用了Vue框架Element Plus前端UI组件来完成。



Process Three

后端设计

本系统的后端为了实现用户权限管理以及前台的临时体验功能，运用了Flask和Django两个框架，Flask框架简约精练，我们用它完成了交互功能，而Django框架则是能更好的完成后端功能中复杂的逻辑调用，为前端提供可以使用的API，更安全地与前端进行信息交流。



关键技术

Content

- 系统流量判别模型
- 网络流量分类模型
- 知识图谱
- 数据处理

more



1. 系统流量判别模型

预训练语言模型（如BERT、LM）在各种NLP任务中取得了显著的成功，为了进一步的使大模型更加适用于我们的任务，在训练模型的过程中，我们运用了深度自注意力蒸馏的方法来压缩基于Transformer的大型预训练模型。运用处理后的模型提取出系统流量数据中的特征，在通过特征工程和K最临近（K-NearestNeighbor，KNN）算法判别出此流量是否为危险流量。





1.1 知识蒸馏

知识蒸馏为深度学习领域新兴的模型压缩及加速技术。针对特定任务的知识蒸馏不同，与任务无关的LM蒸馏模仿原始预训练LM的行为，学生模型可以直接在下游任务上进行调整。我们使用基于任务识别的Transformer LM深度自我注意力提炼框架。其主要思想是深度模仿自我注意模块，这些模块是基于Transformer的教师和学生模型中最重要的组成部分，即将教师网络丰富的知识迁移至学生网络以实现模型压缩并尽量保存模型性能。我们对教师模型最后一次Transformer的自我注意模块进行提炼。

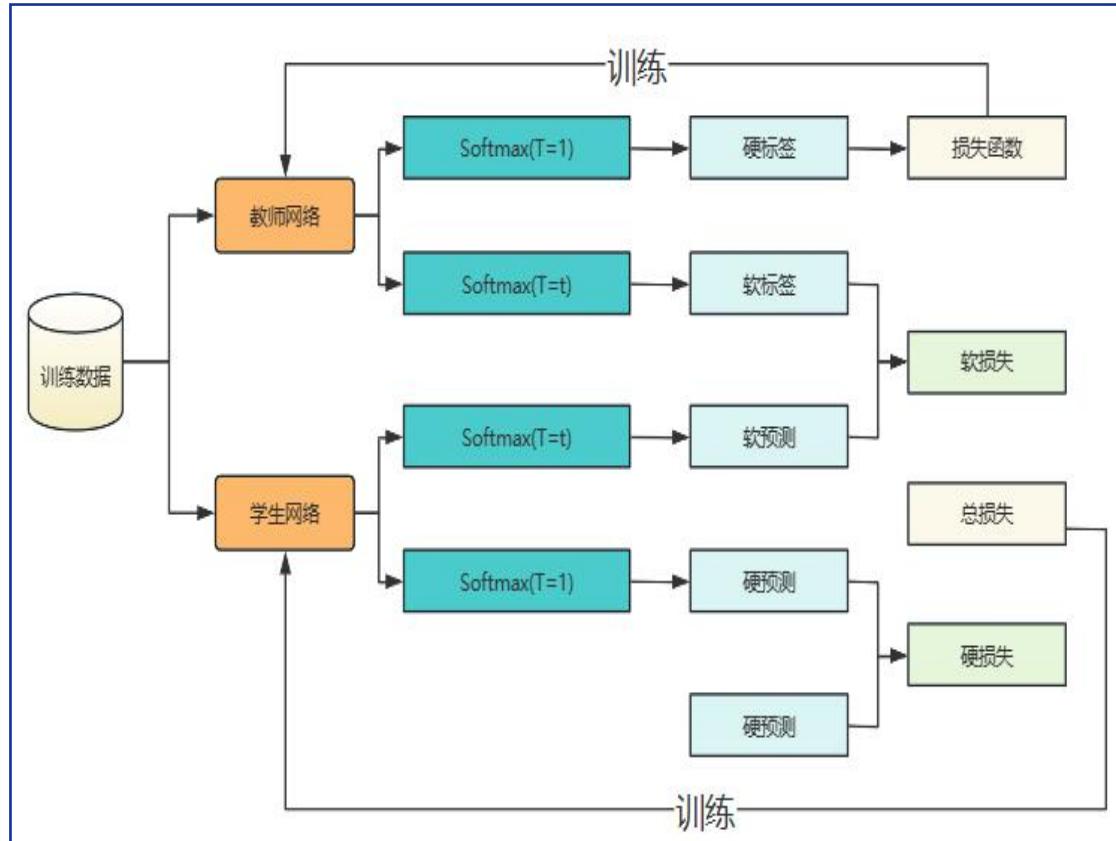


图 1-1 蒸馏过程

蒸馏

模型轻量
化过程

SFDM预测
流程



1.2 模型轻量化过程

通过输入的系统流量序列，使用Sentence-Transformer特征工程提取出特征向量，再将相应的标记嵌入、位置嵌入和段嵌入相加。计算出向量表示 $\{X_i\}_{i=1}^{|X|}$ 。计算出相应的标记嵌入、位置嵌入和段嵌入。

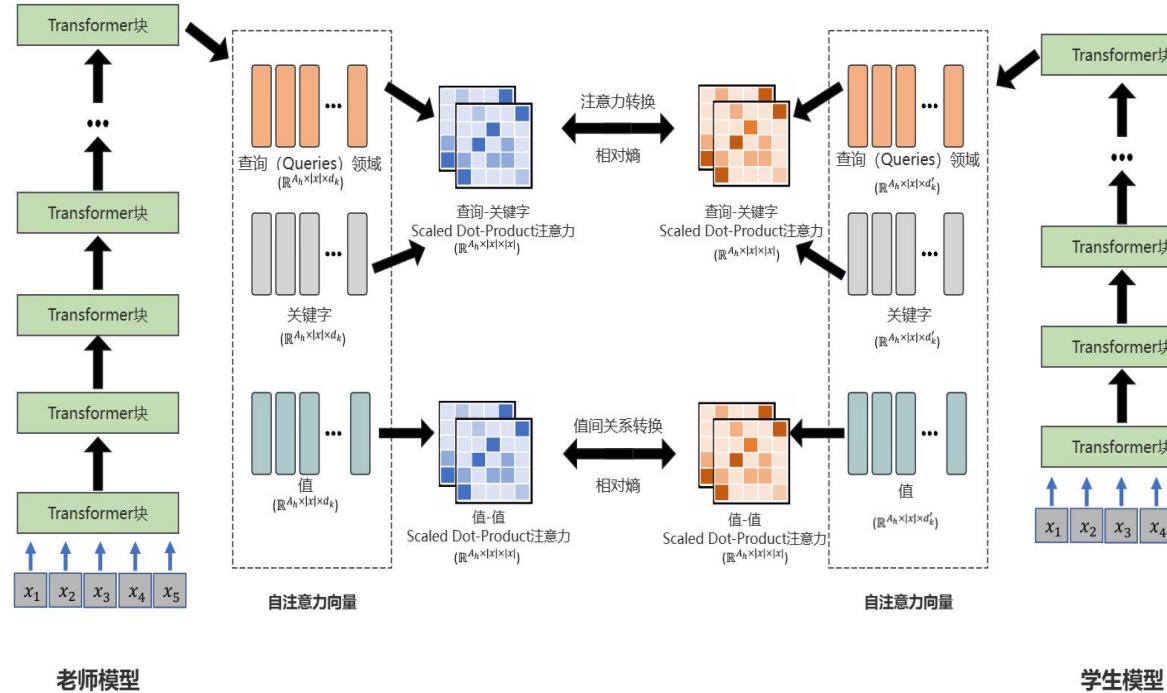


图 1-2 模型轻量化过程

蒸馏

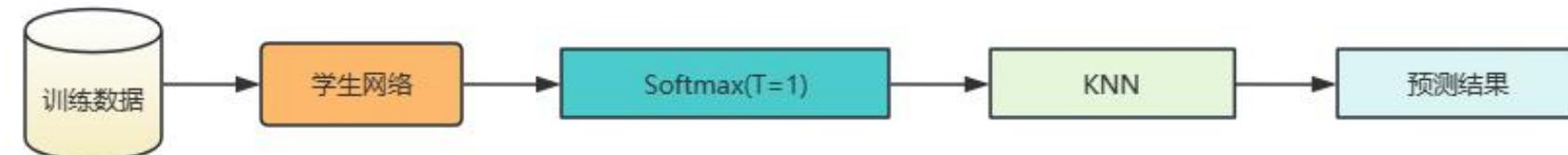
模型轻量化
过程

SFDM预测
流程



1. 3 SFDM预测流程

通过输入的系统流量序列，使用Sentence-Transformer特征工程提取出特征向量，再将相应的标记嵌入、位置嵌入和段嵌入相加。计算出向量表示 $\{X_i\}_{i=1}^{|X|}$ 。计算出相应的标记嵌入、位置嵌入和段嵌入。



蒸馏

模型轻量
化过程

SFDM预测
流程

2. 网络流量分类模型

本文使用的长短期记忆网络 (Long Short-Term Memory Network, LSTM)，相较于传统机器学习方法降低了对于特征工程的依赖，且无需特意去了解专业领域的先验知识即可完成对网络流量数据特征的有效提取，是以能够满足网络流量异常检测及分类的准确性以及鲁棒性。





输出门——LSTM神经元

在图中, f_t 为遗忘门的输出信号, 表示记忆单元c中的遗忘比例。 i_t 为输入门的输出信号, 表示当前输入信息在c中的输入比例。 \tilde{c}_t 为将要输入到c中的预备信息, 与 i_t 进行点乘, 得到c中的信息。 O_t 为输入门的输出信号, 表示c输出到当前状态s中右的比例。为将要输出到隐含层状态h的预备信息, 与 O_t 进行点乘, 得到h中的信息。在时刻t, c_t 经过遗忘门、输入门和输出门的筛选, 得到 h_t 。

图 2-2 NFCM流程图

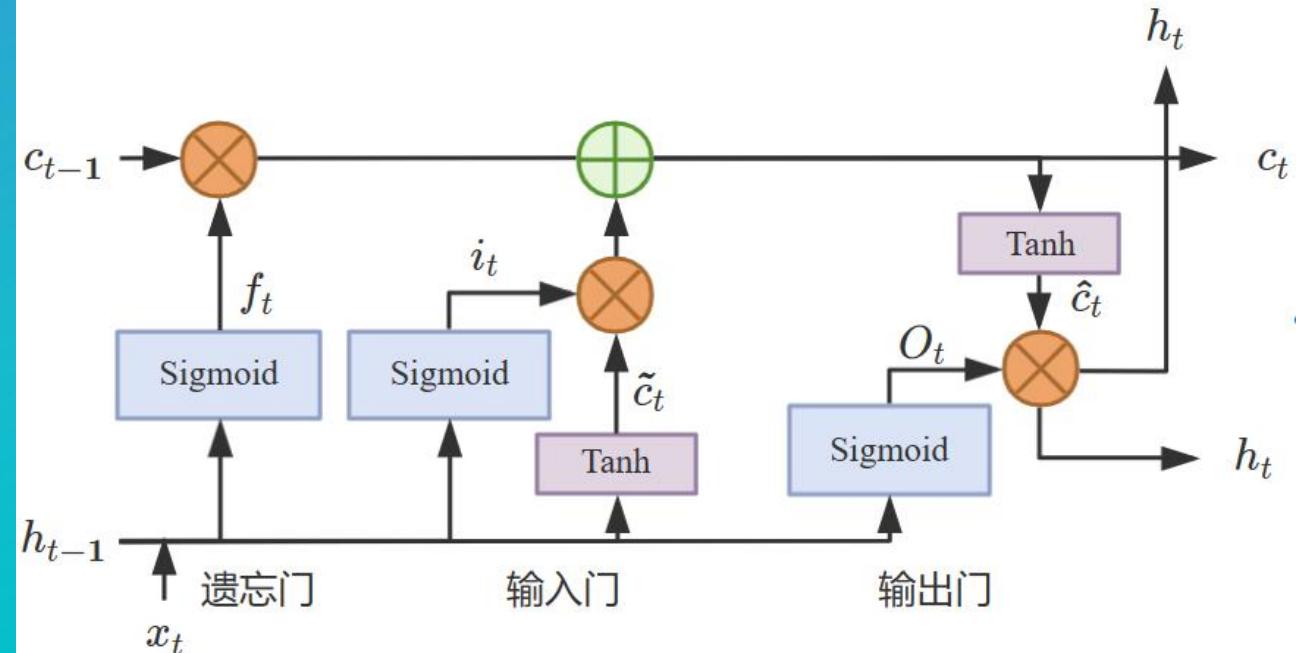
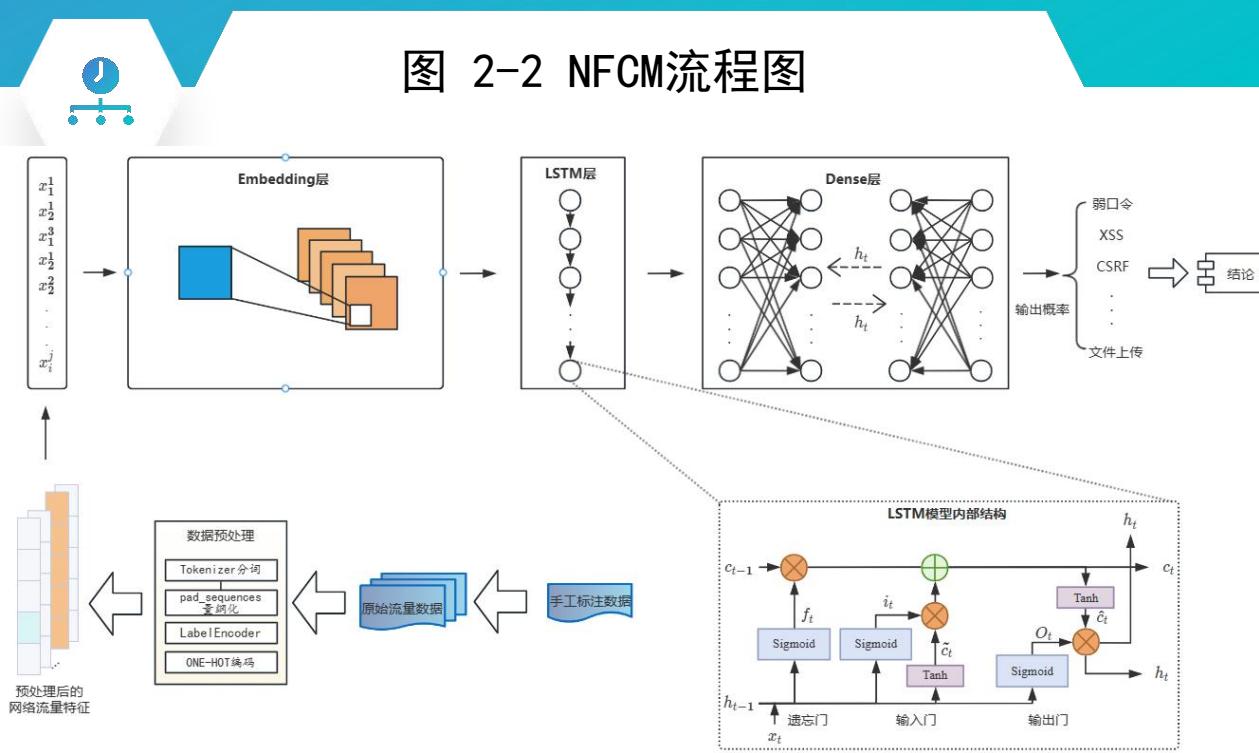


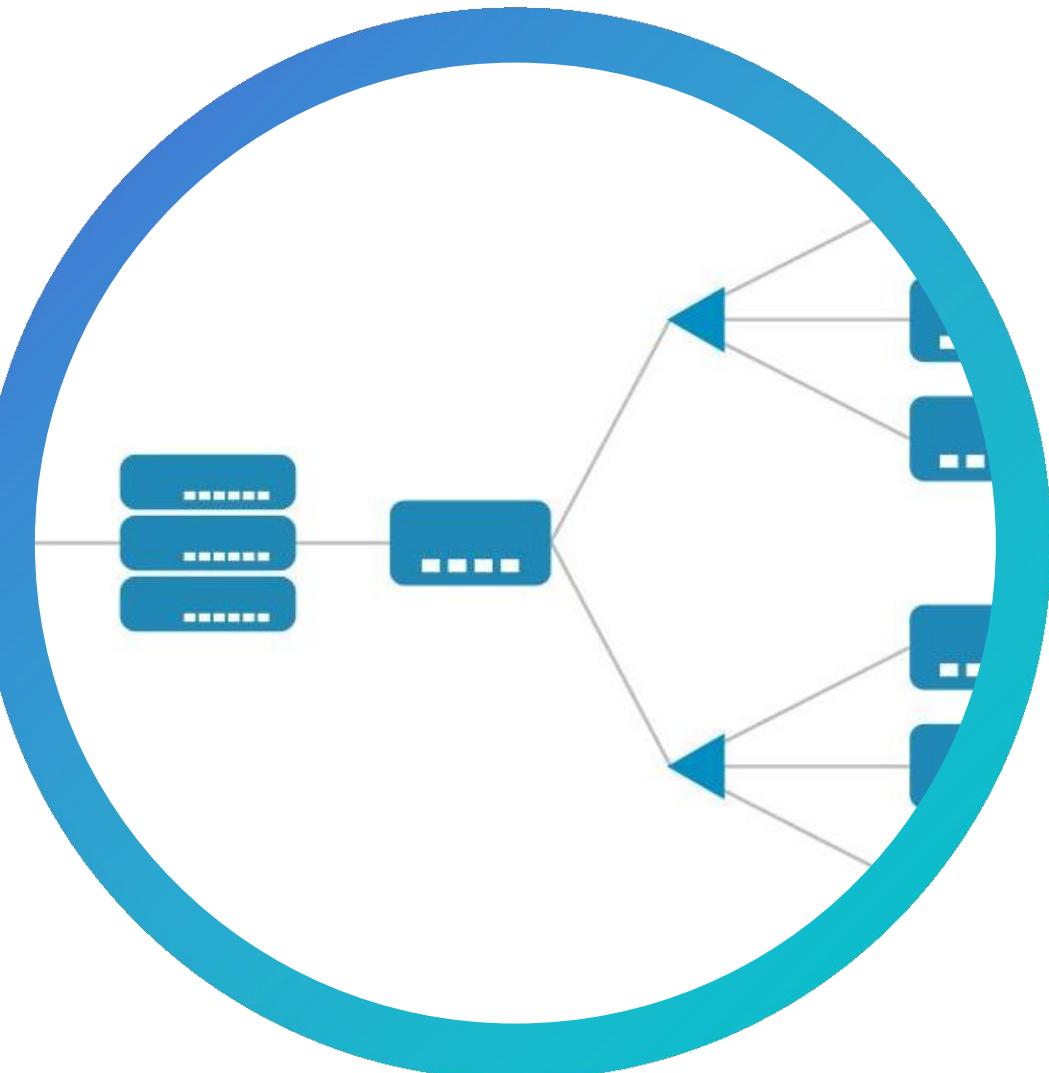
图 2-1 LSTM神经元

NFCM流程图

最后, 一个全连接层 (Dense) 使用 softmax激活函数输出每个类别的预测概率, 输出层的神经元数量等于标签的类别数量, 最后取概率最高的输出为NFCM预测的结果。



3. 网络安全知识图谱——Sherlock之眼



01

数据整合

从多源异构数据中提取关键信息，如安全日志、漏洞数据库、攻击行为等，并进行数据清洗和融合。

02

实体识别

识别数据中的实体，如攻击者、攻击目标、攻击工具等，并构建实体间的关系。

03

知识推理

利用推理算法，发现实体间的新关系，丰富知识图谱的内容。

3. 1 实体识别与关系抽取技术

命名实体识别

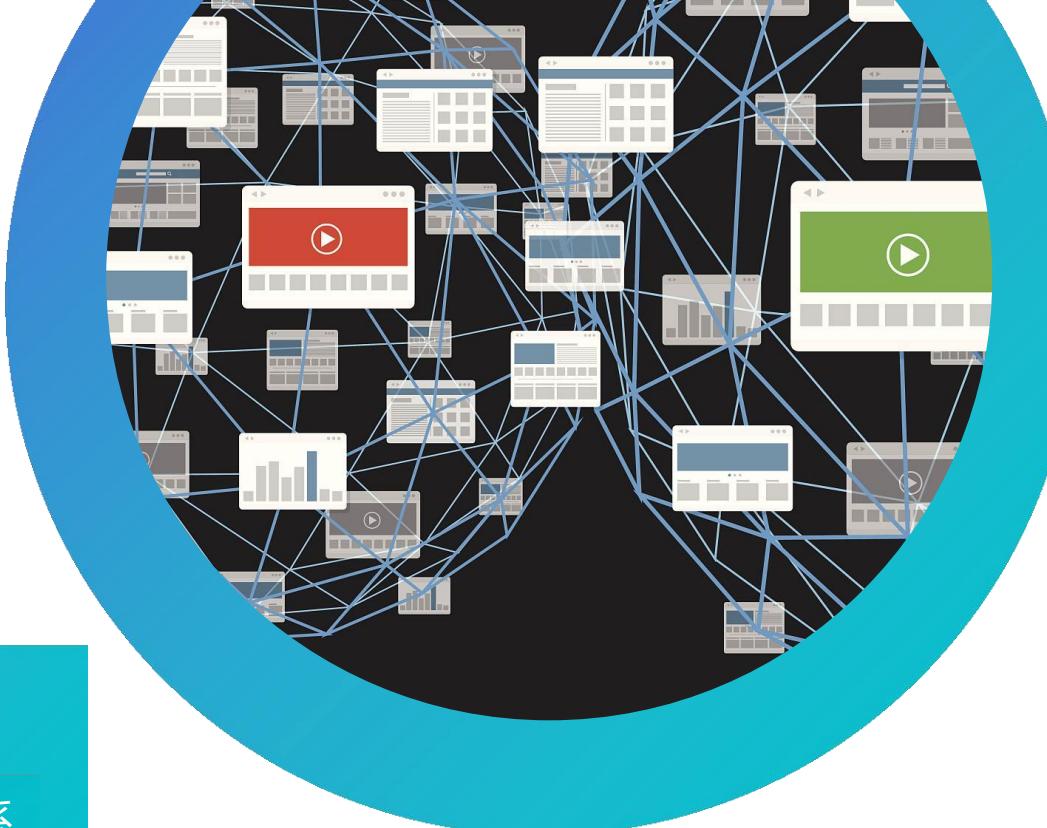
采用自然语言处理等技术，识别文本中的命名实体，如人名、组织名、技术术语等。

关系抽取

分析文本中实体间的语义关系，如攻击者与攻击目标之间的关系、漏洞与攻击工具之间的关系等。

实体链接

将识别出的实体链接到知识图谱中的相应节点，实现知识的整合和共享。



3.2 数据处理



以上三个模型训练使用了不同类型、不同形式的数据，但是数据处理的方法往往大同小异。在数据处理阶段，我们主要采用了数据清洗、数据增强、数据平衡和归一化来提高模型的准确度。

More

More



数据清洗

- 去除重复数据
- 去除噪声数据
- 处理缺失数据



数据增强

- 文本转换
- 数据合成
- 图像旋转



数据平衡和数据归一化

- **数据平衡**: 采样方法, 权重调整, 损失函数调整
- **数据归一化**: 线性归一化, 对数归一化, 正则化

界面设计

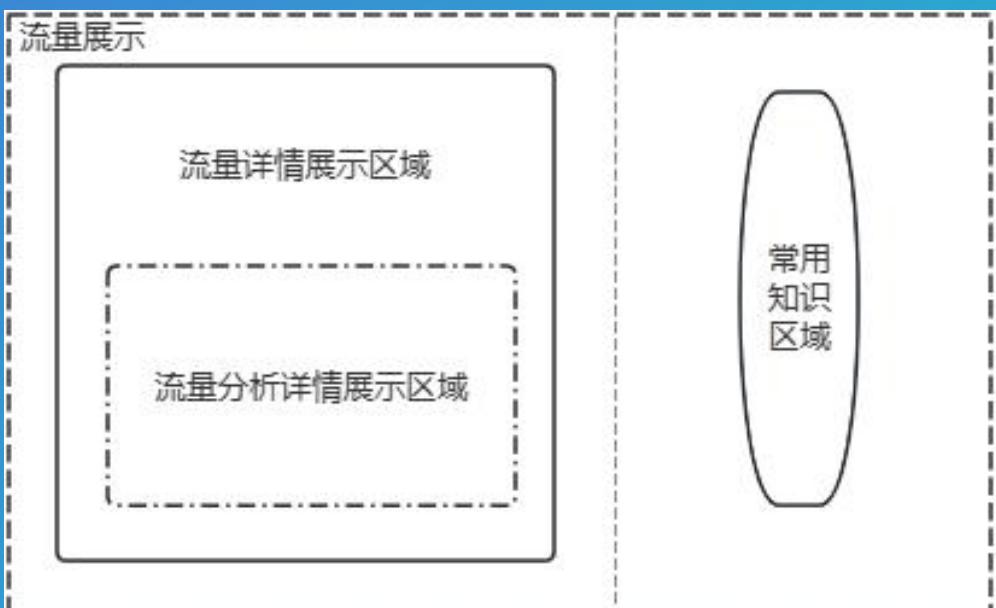
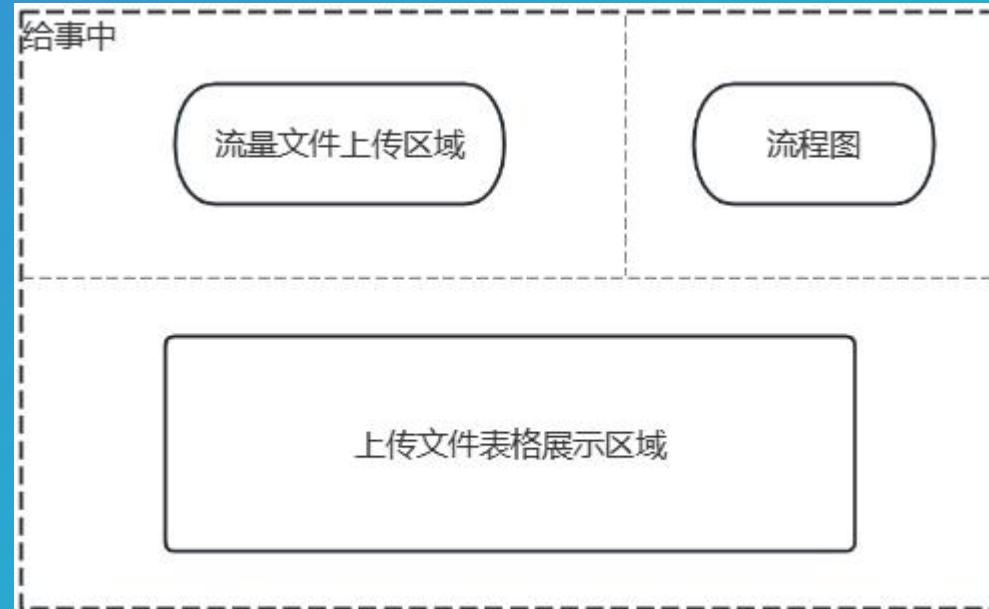
本系统前端页面的实现主要是利用了Vue框架
Element Plus前端UI组件来完成，布局主要
采用了以下几种简约清晰的布局方式

more





前端布局





前台页面

- 为了使人机交互友好，本项目在前台便可以直接使用部分功能，便于用户更快速的适应项目的使用，体验之后我们推荐登录后，开启本系统的完整功能，如右图所示。

The screenshot shows a web browser window with the following details:

- Header:** Includes the logo "流智图灵" (LiuZhiTuLing), the URL "中国大学生计算机设计大赛 – jzjds.bicu.edu.cn", a search bar, and a "首页" (Home) link.
- Right Sidebar:** A notification box displays a green checkmark and the text "检测成功" (Detection successful) followed by "检测结果为: SQLInjection" (Detection result: SQLInjection).
- Main Content Area:** Features a logo and the slogan "基于深度学习的威胁感知平台" (Threat Perception Platform based on Deep Learning). Below this are two tabs: "上传文件分析" (Upload File Analysis) and "数据包分析" (Data Packet Analysis), with "数据包分析" currently selected.
- Text Input:** A large text area contains a POST request payload:

```
POST /scgi-bin/platform.cgi thispage index.htm USERDBUsers.UserName NjVI USERDBUsers.Password  
USERDBDomains.Domainname geardomain '' 5434 '% 3d' 5435 '' MwLJ '% 3d' MwLJ button.login.USERDBUsers.  
router_status Login.Login.userAgent MDpd
```
- Buttons:** A blue "分析" (Analyze) button is located at the bottom of the text input area.
- Footer:** Includes links to "大赛主页" (Competition Home), "系统简介" (System Introduction), "用户协议" (User Agreement), and "联系我们" (Contact Us).

前台页面



给事中

给事中部分总共有系统流量分析和网络流量分析两个功能，顶部的两个按钮都绑定了Ajax事件，用于选择想使用的效果。然后下面上传区域可以选择点击或者拖拽上传，下方的表格会将并解析成功的文件呈现出来

系统流量展示

系统流量列表

| ID | 时间 | 文件名 | 流量类型 | 状态 |
|----|-------------------------------|-------------|--------|-----|
| 1 | 2023-11-12 13:30:57.058843677 | system.pcap | Attack | 待处置 |

相关情报

| | | | | | |
|----------------|-------------------------|---------------|-----------|------------|--------|
| 文件名(File) | system.pcap | 请求方式(Request) | read | 攻击类型(Type) | Attack |
| 异动目录(Abnormal) | /usr/local/sbin/php-fpm | 攻击者(rip) | 127.0.0.1 | | |

攻击行为画像

```

    graph LR
        A[攻击发起] --> B[漏洞利用]
        B -- 攻击路径: 127.0.0.1:45942->127.0.0.1:9000 --> C[提取权限]
        C -- 方法: read --> D[攻击成功]
    
```

攻击路径: 127.0.0.1:45942->127.0.0.1:9000 /usr/local/sbin/php-fpm

系统流量

网络流量

点击或拖拽
即可上传流量文件进行批量上传

- Step 1 选择需要审判的流量类型
- Step 2 按照分类，上传需要研判的流量文件
- Step 3 系统自动进行审计

| 创建日期 | 文件名称 | 流量类型 | 文件大小 | 解析状态 |
|----------------------------|-------------|------|------|------|
| 2024-04-19 14:17:03.839502 | XXE.pcap | 网络流量 | 449 | 解析成功 |
| 2024-04-17 10:06:47.575443 | system.pcap | 系统流量 | 665 | 解析成功 |
| 2024-04-17 10:06:42.365176 | XXE.pcap | 网络流量 | 449 | 解析成功 |

给事中



系统流量展示

后上传解析成功的流量，会在流量审计系统-流量列表中展示出来

常用工具

- 火绒剑、D盾、河马、微步云沙箱

应急响应流程

- 发现异常
- 借助流量设备，查看主机的CPU占用率，出现异常外连端口、主动向外网发起大量连接
- 病毒分析
- 将病毒文件上传到样本库进行分析，对病毒文件进行逆向分析，描绘攻击者画像
- 病毒处理
- 找到病毒对应的进程，删除出病毒文件
- 总结报告
- 对此次事件进行分析，复现总结这次攻击的原因，提出整改意见，输出报告



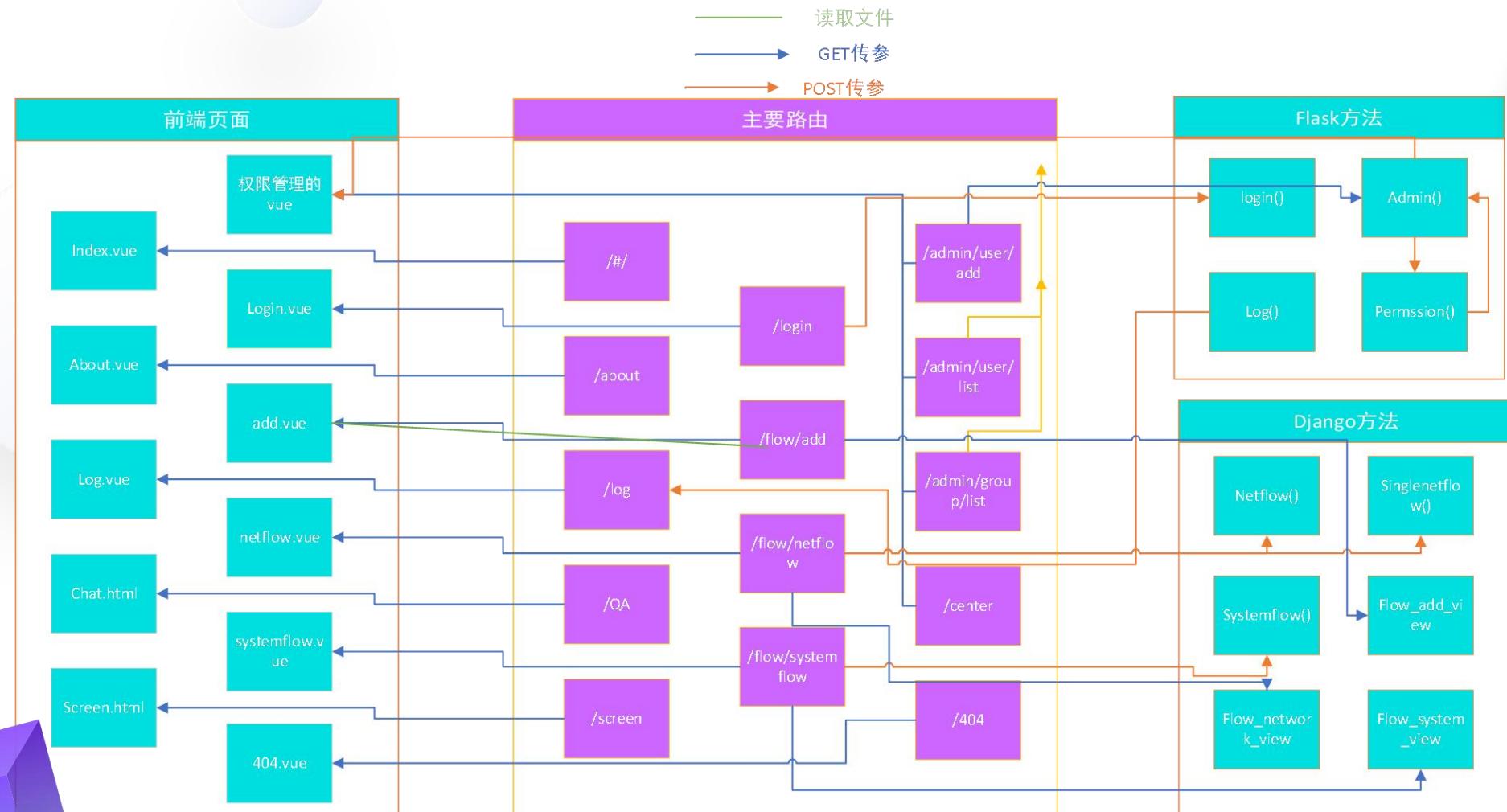
后端设计

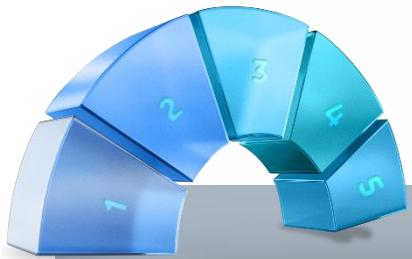
About Us

本系统的后端为了实现用户权限管理以及前台的临时体验功能，运用了Flask和Django两个框架，Flask框架简约精练，我们用它完成了交互功能，而Django框架则是能更好的完成后端功能中复杂的逻辑调用，为前端提供可以使用的API，更安全地与前端进行信息交流。



路由映射图





系统中所有功能模块都是调用自己的接口与方法，给事中模块中，同一个上传控件实现不同类型上传的核心代码如下：

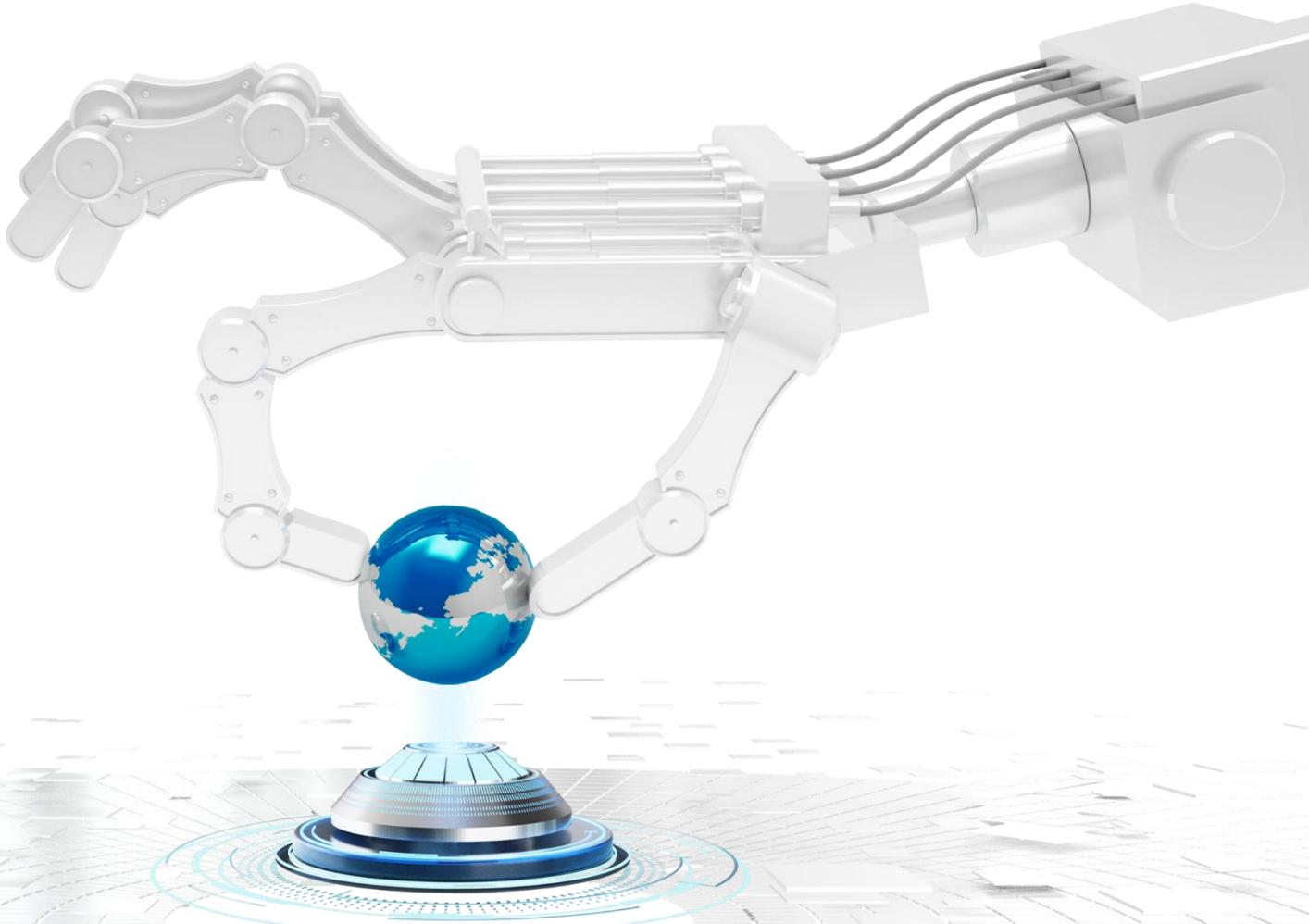
```
<template>
<div style="text-align: center; margin-top:10px">
    <el-radio-group v-model="radioValue" style="display: inline-block;">
        <el-tooltip effect="light" content="<span><strong>
            <a href='https://www.baidu.com' target='_blank' class='fancy-link'>
                下载系统流量样式
            </a>
            </strong></span>" raw-content>
            <el-radio :value="1" size="large" border>系统流量</el-radio>
        </el-tooltip>
        <el-tooltip effect="light" content="<span><strong>
            <a href='https://www.baidu.com' target='_blank' class='fancy-link'>
                下载网络流量样式
            </a>
            </strong></span>" raw-content>
            <el-radio :value="2" size="large" border>网络流量</el-radio>
        </el-tooltip>
    </el-radio-group>
</div>
</template>
<script>
Export default {
    computed: {
        uploadAction() {
            return this.radioValue === 1 ? 'http://localhost:8000/api/systemflowpredictor/'
                : 'http://localhost:8000/api/netflowpredictor/'
        },
    },
}
</script>
```



PART FIVE

项目总结

在本系统的开发过程中，我们团队经历了一系列的挑战和机遇。通过这个项目，我们掌握了如何进行项目协调、任务分解、克服困难、水平提升、升级演进和商业推广等诸多方面的知识和技能。



项目总结

Project summary

首先，我们的项目协调和任务分解非常重要，这对于项目的进展和成果至关重要。我们需要清晰地定义项目的范围、目标和时间表，并根据团队成员的技能和能力来分配任务，将其转化为可执行的任务列表。通过这个过程，我们可以确定每个团队成员的责任和角色，并确保每个人都明确自己的任务和工作计划。我们还使用了项目管理工具——Teambition，以确保项目按计划顺利进行。

其次，我们克服了一些技术上的困难，特别是在研究和训练深度学习的模型时。这需要团队成员不断学习和探索新技术，同时也需要投入大量时间和精力。但最终，我们成功地训练了可以完成网络流量分类和系统流量判别的模型，并实现了Sherlock智能问答以及其他的功能。

同时，通过这个项目，我们的团队也获得了极大的水平提升。我们不仅学会了如何使用自然语言处理算法和其他相关技术，而且还提高了我们的团队协作和沟通能力。在这个过程中，我们的团队成员更好地理解了彼此的技能和能力，并协同工作以实现最佳效果。

最后，我们计划将项目进行升级演进和商业推广。在未来，我们将继续改进和完善流智图灵的功能，包括增加新的特性和服务，并且我们将寻求商业合作伙伴，以将这个项目推向更广泛的市场。总之，流智图灵的开发是一个挑战和机遇并存的过程。通过项目协调、任务分解、克服困难、水平提升、升级演进和商业推广等方面的努力，我们取得了成功，并且为未来的发展打下了坚实的基础。

About Us



THANK YOU.

Flowwise Turings

Contact

About Us

成员/Team member

指导老师/advisor

