ICS［点击此处添加 ICS 号］
CCS ［点击此处添加中国标准文献分类号］

# 团 体 标 准

T/BISA xxx-2026

# 基于人工智能的工业协议解析技术应用指南

# Application Guidance of Artificial Intelligence Based Industrial Protocol Analysis Technology

（征求意见稿）

2026－XX－XX 发布　　2026－XX－XX 实施

北京软件和细心你服务业协会　　发布

# 目　　次

# 前　言

本文件按照GB/T 1.1—2020《标准化工作导则　第1部分：标准化文件的结构和起草规则》的规定起草。

本文件由北京信联数安科技有限公司提出，由北京软件和信息服务业协会归口。

本文件起草单位：北京信联数安科技有限公司、北京邮电大学、信联科技（南京）有限公司、北京明博信安信息技术有限公司、华夏宏源（北京）科技有限公司、工信君阳（北京）科技有限公司、北京软件和信息服务业协会

本文件主要起草人：xx、xx、xx。

本文件于2026年X月首次发布。

# 基于人工智能的工业协议解析技术应用指南

## 1 范围

本文件给出了基于人工智能的工业协议解析技术的总体原则、应用框架以及面向协议规范、网络流量、二进制程序进行协议解析的应用指南。

本文件适用于指导工业协议智能分析相关系统或功能模块的设计、开发与测试。

## 2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB/T 36323-2018 信息安全技术—工业控制系统安全管理基本要求

GB/T 45288.1-2025 人工智能 大模型 第1部分:通用要求

## 3 术语和定义

GB/T 36323-2018、GB/T 45288.1-2025界定的以及下列术语和定义适用于本文件。

3.1

**工业协议规范 Industry Protocol Specification**

协议规范是指对工业通信协议的核心逻辑、交互规则、格式要求、技术参数等进行统一界定和明确描述的规范性文件，是工业设备间、系统间实现数据交互、通信兼容的依据。

3.2

**协议解析 Protocol Analysis**

协议解析是指在不依赖协议规范的情况下，通过对协议实体的网络输入输出、系统行为和指令执行流程进行监控和分析，提取协议格式以及状态机的过程。

## 4 符号和缩略语

下列缩略语适用于本文件。

| | |
|---|---|
| API | 应用程序编程接口（Application Programming Interface） |
| RFC | 互联网技术标准文档（Request For Comments） |

## 5 总体原则

总体原则要求如下：

a) 应能准确解析出工业通用协议及私有协议的格式、字段及交互流程，误检率和漏检率应满足实际场景需求；

b) 在采集、处理包含工业敏感数据时，应采取严格的数据保护措施，确保数据的安全性和隐私性，防止数据泄露、篡改和丢失。

## 6 应用框架

基于人工智能的工业协议解析技术以多种工业协议载体为输入，利用人工智能算法或大模型对协议载体进行智能解析，实现工业通用协议及私有协议的消息格式、字段定义及交互流程的特征输出。基于人工智能的工业协议解析技术应用框架如下图1所示。

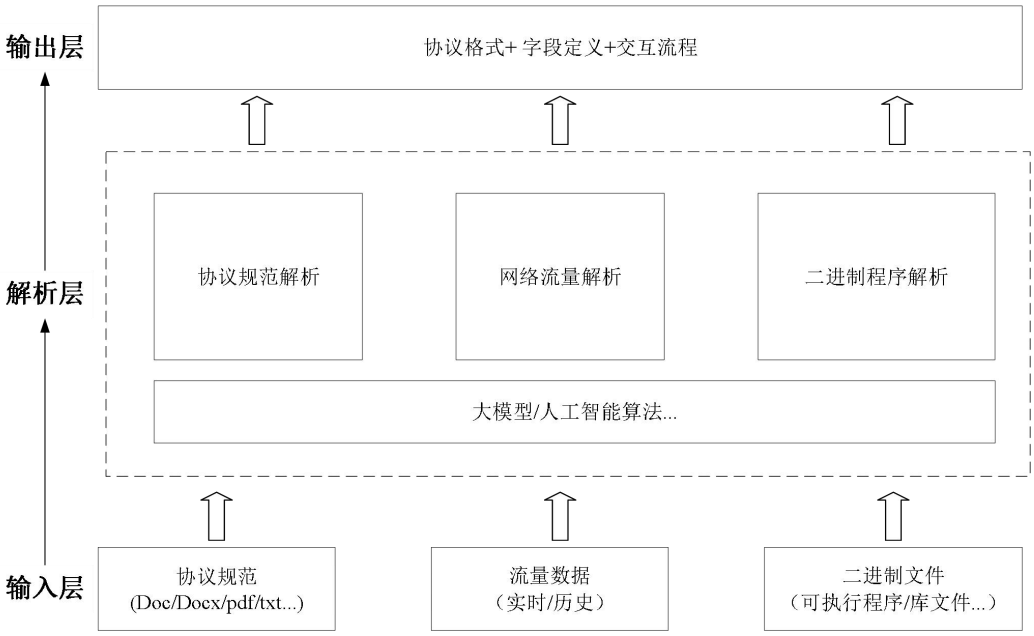


图1 基于人工智能的工业协议解析技术应用框架

基于人工智能的工业协议解析框架包括：

a) 输入层：包括工业协议规范、工业流量数据及工业二进制文件。
   1) 工业协议规范包括通用协议RFC文档、行业标准协议手册以及厂商私有协议说明等；
   2) 工业流量数据包括实时采集的流量文件以及历史流量抓包文件。
   3) 工业二进制文件包括可执行业务程序及通信库文件等
b) 解析层：包括大模型及人工智能算法，以及基于人工智能的协议解析方法，包括针对工业协议规范的解析方法、针对网络流量的解析方法、针对二进制程序的解析方法。
   1) 协议规范解析利用大模型对协议规范文本进行解析，实现从非结构化的协议文档中自动提取出协议相关的结构化信息，包括协议消息格式、协议字段定义、协议交互流程等；
   2) 网络流量解析采用人工智能算法，以网络数据流为分析对象，根据协议字段的取值变化频率和特征推断得到协议格式。
   3) 二进制程序解析利用大模型对目标程序的中间代码进行静态分析，特别是重点分析与网络通信相关函数，如socket调用、数据序列化操作的实现逻辑，来推断通信协议的格式、规则和行为。
c) 输出层：对各种协议载体进行解析后得到的工业协议特征，包括协议消息格式、字段定义、交互流程等。
   1) 协议消息格式包括报文头部、报文体、报文尾部的具体组成形式，其中报文头部可包含起始标识、版本号、报文长度、优先级、校验码等控制字段；报文体可包含业务数据字段、指令字段等核心内容；报文尾部可包含结束标识、校验字段等；

2) 字段定义包括字段名称、字段标识、数据类型、字段长度、取值范围、字段含义、约束条件及字段之间的关联关系以及私有协议中自定义字段的特殊规则；
3) 交互流程包括会话建立、数据交互、会话终止的全流程步骤，以及各步骤的触发条件、执行顺序及交互时序。

## 7 协议规范解析

### 7.1 文本清洗

文本清洗功能要求如下：

a) 应过滤并去除协议规范文本中的版权声明、版本历史、参考文献等与协议核心逻辑无关的文本片段；
b) 宜剔除文本中的无效字符，包括但不限于乱码、特殊符号、无意义占位符、冗余空格、换行符及不可打印字符，确保文本格式规整；
c) 应不改变协议核心逻辑、关键字段定义及参数取值，确保清洗后文本的准确性和完整性，为后续大模型解析提供可靠的文本基础；
d) 可支持自定义清洗规则，根据不同类型工业协议的文本特点，灵活配置需过滤、保留的内容，适配多场景协议文本清洗需求。

### 7.2 结构化分块

结构化分块功能要求如下：

a) 应基于标题层级将长文本拆分为消息格式块，宜基于语义特征将长文本拆分为交互过程块，便于大模型聚焦处理。
b) 消息格式块应聚焦协议消息的核心要素，包括但不限于字段定义、字段长度、数据类型、编码规则、默认取值、约束条件等，形成结构化的消息描述单元；
c) 交互过程块应梳理协议的通信逻辑，包括但不限于请求-响应机制、时序关系、交互步骤、异常处理流程、会话建立与终止规则等，明确各交互环节的关联关系；
d) 可根据协议类型自定义分块规则，灵活配置分块粒度、标签体系，适配不同工业协议的结构化解析需求。

### 7.3 数据标注

若采用大模型微调方法，数据标注功能要求如下：

a) 样本来源应真实可靠，优先选用工业现场实际产生的协议数据、官方协议规范配套示例数据，避免人工虚构数据，确保样本与实际应用场景高度贴合；
b) 样本宜覆盖目标解析工业协议的主要类型、核心场景及关键字段；
c) 对消息格式样本应标注字段名、数据类型、长度、约束条件、字段间依赖关系等；
d) 对交互过程样本应标注参与方、消息触发条件、时序顺序、状态转换规则、异常分支等。

### 7.4 协议规约提取

协议规约提取功能要求如下：

a) 应利用大模型提取出协议消息的字段组成、结构约束、编码规则等信息；
4) 应支持从协议规范文本中提取独立的消息单元或功能码，如心跳、采集、控制、错误等消息类型特征；
5) 应支持从消息类型特征字段的核心属性的提取，如字段名、数据类型、长度、默认值、取值范围等，并识别字段的层级结构；

6) 应支持字段依赖关系、校验规则等隐含信息的提取。

b) 应利用大模型还原出协议参与方的消息收发时序、状态转换、异常处理逻辑等信息。
   1) 应能够识别交互参与方，确定协议的通信主体；
   2) 应能够提取消息交互序列，按时间顺序梳理消息的收发逻辑；
   3) 应能够建模状态转换与异常分支，提取协议的状态机逻辑及异常处理流程。

### 7.5 结果验证

结果验证功能要求如下：

a) 应具备人工校验机制，对模型提取的消息格式字段完整性、交互流程逻辑一致性进行抽样检查，重点验证隐含约束和异常分支的准确性。
b) 宜具备自动化校验机制，利用大模型将提取的消息格式转换为协议解析代码，输入真实协议数据包，验证解析结果是否与预期一致。

## 8 网络流量解析

### 8.1 流量去噪与过滤

流量去噪与过滤功能要求如下：

a) 应剔除工业网络中的无关流量，包括但不限于广播包、组播包、已知非目标协议流量、无效测试包及干扰报文，仅保留未知协议流量和目标解析协议的数据流；
b) 应按网络五元组对保留的数据流进行会话重组，还原完整的协议交互会话，形成连续、完整的报文序列，并提取协议应用层载荷；
c) 应对变长报文按固定头部长度或特征字节进行对齐，消除因载荷长度差异导致的序列偏移。

### 8.2 报文聚类

报文聚类功能要求如下：

a) 应从载荷中提取初始特征，如报文长度、特定字节取值、字节频率分布等；
b) 应根据特征相似性对报文进行聚类；
c) 应对聚类后的报文簇进行字段边界分析，推断字段的起始、结束位置与长度，将连续的载荷拆分为独立的字段。

### 8.3 协议规约提取

协议规约提取功能要求如下：

a) 确定字段边界后，应进一步分析每个字段的语义含义，如序列号、长度、校验和以及约束规则，如取值范围、依赖关系等；
   1) 应支持字段语义推断，通过统计分布、上下文位置以及与已知协议字段的相似性推测字段可能的含义，如长度字段、命令字、状态位、校验和等；
   2) 应支持报文格式生成，通过将单个簇的字段信息整合，抽象为通用的报文格式规范。为每个报文簇定义报文类型，按顺序描述每个字段的属性，如字段名、偏移位置、长度、类型、取值约束、编码方式、语义说明等。
b) 应支持基于单个报文格式，进一步分析报文序列的时序关系，还原协议的交互流程与状态机逻辑；
   1) 应支持交互序列提取，按时间顺序梳理通信双方的报文收发序列，识别请求-响应成对关系；
   2) 应支持交互阶段划分，如连接建立、认证、数据传输、心跳保活、连接断开等；
   3) 应支持状态机建模，能够提取协议状态、状态转换规则及异常处理逻辑。

### 8.4 结果验证

结果验证功能要求如下：

a) 应具备人工校验机制，对模型提取的消息格式字段完整性、交互流程逻辑一致性进行抽样检查，重点验证隐含约束和异常分支的准确性。
b) 宜具备自动化校验机制，利用大模型将提取的消息格式转换为协议解析代码，输入真实协议数据包，验证解析结果是否与预期一致。

## 9 二进制程序解析

### 9.1 程序汇编还原

程序汇编还原功能要求如下：

a) 应对变量名、结构体定义等进行重命名或注释补全，以增强可读性；
b) 应将中间代码切分成适合大模型上下文窗口长度的片段或单元；
c) 并保留与网络通信相关的部分，如socket API调用、加密/解密函数、数据打包/解包逻辑等。

### 9.2 网络通信函数定位与识别

网络通信函数定位与识别功能要求如下：

a) 应支持识别操作系统提供的网络编程接口，以及加密库或序列化库的调用；
b) 应支持通过分析函数的参数传递和返回值，并结合交叉引用追踪调用关系，定位所有网络通信入口点；
c) 应支持追踪send类函数的数据缓冲区来源，并通过标记每个操作的偏移量与操作类型，勾勒协议头的格式；
d) 应支持追踪recv类函数的数据缓冲区去向，并通过分析函数对缓冲区的拆分方法，还原原始协议数据。

### 9.3 协议规约提取

协议规约提取功能要求如下：

a) 应支持对还原的原始协议数据进行字段属性提取，包括数据单元的位置、长度、类型、编码方式等；
b) 应支持提取协议的动态行为规则，形成完整的协议规约；
c) 应支持协议状态机构建，分析程序在不同网络状态下的处理逻辑，通过控制流图梳理状态转换条件；
d) 应支持通过分析不同路径下发送/接收的消息序列，构建协议状态机，明确状态转换条件与触发事件。
e) 应支持分析异常分支对应的协议容错机制，解析错误码字段与对应的处理逻辑，如重连、重试、断开连接等。

### 9.4 结果验证

结果验证功能要求如下：

a) 应具备人工校验机制，对模型提取的消息格式字段完整性、交互流程逻辑一致性进行抽样检查，重点验证隐含约束和异常分支的准确性。
b) 宜具备自动化校验机制，利用大模型将提取的消息格式转换为协议解析代码，输入真实协议数据包，验证解析结果是否与预期一致。