ICS 点击此处添加 ICS 号
CCS 点击此处添加 CCS 号

# 团 体 标 准

T/CPUMT XXXX—XXXX

# 工业互联网平台 工业模型与组件通用技术要求

Industrial Internet platform—General requirements for industrial mechanism models and components

（征求意见稿）

（完成时间：2024-5-7）

**在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。**

XXXX - XX - XX 发布 XXXX - XX - XX 实施

中国和平利用军工技术协会 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。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别专利的责任。

T/CPUMT DDDD—XXXX《工业互联网平台 工业模型与组件通用技术要求》与T/CPUMT AAAA—XXXX《工业互联网平台 总体技术要求》、T/CPUMT BBBB—XXXX《工业互联网平台 边缘层通用技术要求》、T/CPUMT CCCC—XXXX《工业互联网平台 工业大数据通用技术要求》、T/CPUMT EEEE—XXXX《工业互联网平台 工业APP通用技术要求》、T/CPUMT FFFF—XXXX《工业互联网平台 服务通用技术要求》、T/CPUMT GGGG—XXXX《工业互联网平台 开发及运行环境通用技术要求》共同构成工业互联网平台的研发、建设及部署的基本型标准体系。

本文件由中国和平利用军工技术协会提出并归口。

本文件起草单位：

本文件主要起草人：

# 引　言

工业互联网是新一代信息通信技术与工业经济深度融合的新型基础设施、应用模式和工业生态，通过对人、机、物、系统等全面连接，构建起覆盖全要素、全产业链、全价值链的全新制造和服务体系，为工业乃至产业数字化、网络化、智能化发展提供了实现途径。在支撑制造强国网络强国建设，提升产业链现代化水平，推动经济高质量发展方面发挥了重要作用。党的二十大报告指出，坚持把发展经济的着力点放在实体经济上，推进新型工业化，推动制造业高端化、智能化、绿色化发展，促进数字经济和实体经济深度融合。2018年为贯彻落实《国务院关于深化“互联网+先进制造业”发展工业互联网的指导意见》，加快推进工业互联网创新发展，加强对有关工作的统筹规划和政策协调，经国家制造强国建设领导小组会议审议，设立国家制造强国建设领导小组工业互联网专项工作组。在《工业互联网专项工作组2022年工作计划》中指出“加快工业互联网基础共性、关键技术、典型应用等产业亟需标准研制”。

工业互联网平台作为工业全要素链接的枢纽与工业资源配置的核心，在工业互联网体系架构中具有至关重要的地位。工业互联网平台是面向制造业数字化、网络化、智能化需求，构建基于海量数据采集、汇聚、分析的服务体系，支撑制造资源泛在连接、弹性供给、高效配置的工业云平台。其本质是通过构建精准实时、高效的数据采集互联体系，建立面向工业大数据存储、集成、访问、分析管理的开发环境，实现工业技术、经验、知识的模型化、标准化、软件化、复用化，不断优化研发设计、生产制造、运营管理等资源配置效率，形成资源富集、多方参与、合作共赢、协同演进的制造业新生态。

《工业互联网平台》系列标准分为7个文件：

——T/CPUMT　AAAA—XXXX　工业互联网平台　总体技术要求

——T/CPUMT　BBBB—XXXX　工业互联网平台　边缘层通用技术要求

——T/CPUMT　CCCC—XXXX　工业互联网平台　工业大数据通用技术要求

——T/CPUMT　DDDD—XXXX　工业互联网平台　工业模型与组件通用技术要求

——T/CPUMT　EEEE—XXXX　工业互联网平台　工业APP通用技术要求

——T/CPUMT　FFFF—XXXX　工业互联网平台　服务通用技术要求

——T/CPUMT　GGGG—XXXX　工业互联网平台　开发及运行环境通用技术要求。

本文件规定了工业模型通用要求、组件通用要求，有利于工业模型及组件的研发企业、应用单位和第三方测评机构开展工业模型的开发、管理、应用工作，推动模型融合创新，助力建设全新的工业模型全生命周期服务体系。

# 工业互联网平台 工业模型与组件通用技术要求

## 1 范围

本文件规定了工业模型及组件的通用要求。

本文件适用于企业、开发者等开展工业模型开发、管理、部署、服务等活动，也适用于应用企业、第三方测评机构开展管理和评价工业模型及组件等活动。

## 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 35295 信息技术大数据 术语

## 3 术语和定义

GB/T 35295界定的以及下列术语和定义适用于本文件。

3.1

**结构化数据 structured data**

一种数据表示形式，按照此种形式，由数据元素汇聚而成的每个记录的结构都是一致的并且可以使用关系模型予以有效描述。

[来源：GB/T 35295—2017，2.2.13]

3.2

**非结构化数据 unstructured data**

数据结构不规则或不完整，没有预定义的数据模型，不方便用数据库二维逻辑表来表现的数据。

注：包括所有格式的办公文档、文本、图片，HTML、各类报表、图像和音频/视频信息等。

3.3

**半结构化数据 semi-structured data**

不符合关系型数据库或其他数据表的形式关联起来的数据模型结构，但包含相关标记。用来分隔语义元素以及对记录和字段进行分层的一种数据化结构形式。

[来源：GB/T 38637.2—2020，3.5]

## 4 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）

AUC：受试者工作特征曲线下面积（Area Under Curve）

CPU：中央处理器（Central Processing Unit）

GPU：图形处理器（Graphic Processing Unit）

JAR：软件包文件格式（Java Archive）

PRC：精确率召回曲线（Precision-Recall Curve）

ROC：接受者工作特征曲线（Receiver Operating Characteristic）

SDK：软件开发工具包（Software Development Kit）

## 5 工业模型通用要求

### 5.1 工业模型概述

工业模型是根据工业生产过程的内部机理，运用行业知识、定理定律和专家经验，结合人工智能算法建立的数学模型，可用于仿真、分析、预测、优化、决策等多种工业场景。

## 5.2 工业模型分类

当前，工业模型存在多种分类维度，见表1。

表1 工业模型类型

| 序号 | 分类维度 | 模型类型 |
|---|---|---|
| 1 | 模型可解释性 | 工业机理模型（白箱模型）、工业智能模型（黑箱模型） |
| 2 | 应用行业 | 航空/航天、石油化工、能源、电力、电子、汽车、材料、机械等 |
| 3 | 专业学科 | 力学、电磁学、热力学、化学、声学等 |
| 4 | 应用场景 | 研发设计、生产制造、质量管控、仓储物流、安全生产、节能减排、供应链管理、运营管理、运维服务 |
| 5 | 目标对象 | 部件级、设备级、产线级、车间级、工厂级、跨工厂级、企业级、跨企业级 |
| 6 | 使用方式 | 接口类、文件类 |
| 7 | 模型功能 | 仿真类、分析类、预测类、优化类、决策类等 |
| 8 | 模型交互 | 独立模型、交互式模型 |

## 5.3 工业模型全生命周期

### 5.3.1 概述

工业模型项目通常以需求、数据、代码、算法为输入，以模型、模型服务为输出。其生命周期主要由需求设计、模型开发、模型管理、模型服务4个阶段组成，涵盖需求管理、数据处理、模型开发、模型管理和部署运营等过程，工业模型全生命周期见图1。

a) 需求管理：根据商业目标与业务需求，开展可行性分析，编制技术需求和技术方案等规范性文件，指导后续模型开发阶段工作的开展。
b) 数据处理：将源数据进行处理，获取适用于模型训练的数据。
c) 模型开发：在实验环境中，对模型进行训练、调优、评估选择等过程，得到最优模型。
d) 模型管理：对于自训练、第三方工业模型及模型资产进行统一纳管和测试。
e) 部署运营：将模型、配置等封装后部署至目标环境，并对已在生产环境上线的模型服务提供维护和更新能力。

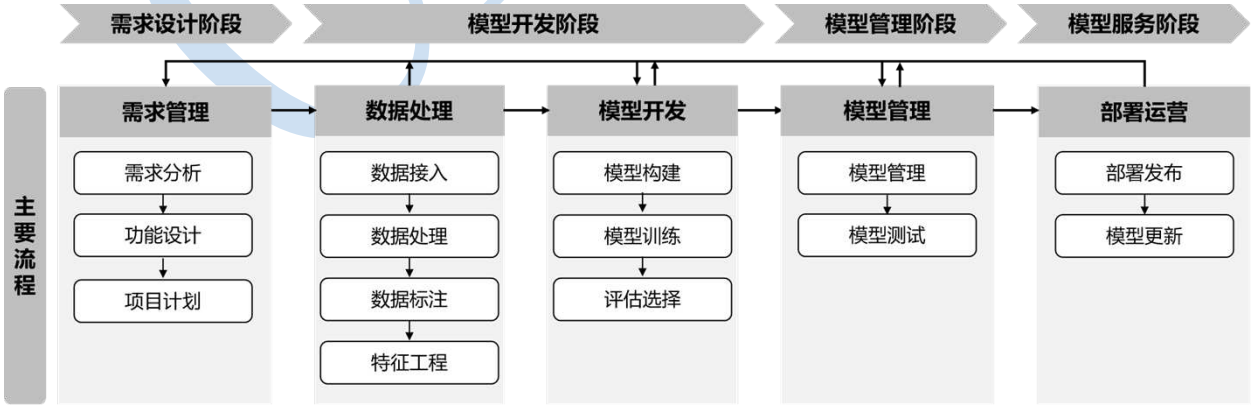



图1 工业模型全生命周期

### 5.3.2 需求设计阶段

需求设计阶段的主要活动是对模型需求方的业务需求进行分析和设计，以解决工业模型项目存在的需求管理流程混乱、需求理解不一致、风险不可控等问题，旨在从源头提升项目质量，降低需求变更带来的影响。该阶段主要包含需求分析、功能设计、项目计划等环节。工业模型需求设计阶段标准流程参考附录A。

a) 主要输入：业务需求。
b) 主要步骤：
   1) 将业务需求转化为技术问题，评估使用工业模型解决业务问题的可行性、必要性和潜在风险；
   2) 设计工业模型项目架构，完成技术选型；
   3) 梳理项目过程的数据源、数据处理规则需求，并根据后续的反馈持续迭代更新。
c) 主要输出：需求分析报告、功能设计文档、项目计划。

### 5.3.3 工业模型开发阶段

工业模型开发阶段的主要活动是通过对源数据进行接入、处理、存储、分析等操作，形成可用于模型开发与服务的高质量数据集，进而基于选定的算法，完成模型的训练、功能测试、评估和选择。该阶段主要包含数据接入、数据处理、数据标注、特征工程、模型构建、模型训练、评估选择等环节。

a) 主要输入：源数据、项目架构。
b) 主要步骤：
   1) 接入并提取源数据；
   2) 分析数据质量，根据需求有选择地开展数据处理、标注等工作，形成模型训练数据集；
   3) 选取适宜的工具，进行模型的开发与训练；
   4) 对模型进行功能测试，结合评估指标，综合优化模型的结构、参数等因素，直至获得最优模型。
c) 主要输出：处理后的数据/特征、算法脚本、模型文件、评价指标。

### 5.3.4 工业模型管理阶段

工业模型管理阶段的主要活动是基于标准化接口对自训练模型和第三方模型进行统一纳管，并对包括模型文件、使用文档、配置文件等在内的模型资产进行集中盘点管理；通过对模型进行非功能测试，确保模型具备部署发布的能力。该阶段主要包含模型管理、模型测试等环节。

a) 主要输入：模型文件、环境/参数配置。
b) 主要步骤：
   1) 将模型文件、参数配置、标签、使用文档等文件通过标准化接口上传至模型仓库进行统一集中管理，模型上传信息规范参考附录 B；
   2) 将代码、模型、配置等要素进行构建打包和集成测试，产出支持交付的部署包，包括镜像文件、JAR 包等；
   3) 进行沙箱验证、AB 实验等非功能测试，完成模型服务上线验证。
c) 主要输出：已纳管的工业模型、模型测试结果。

### 5.3.5 工业模型服务阶段

工业模型服务阶段的主要活动是对已纳管且完成测试的工业模型配置和管理依赖参数，形成模型服务、部署至目标环境并对模型及服务情况进行监控，实现模型的迭代更新。该阶段主要包含部署发布、模型更新等环节。

a) 主要输入：模型部署需求、监控指标数据、触发条件、新数据/特征。
b) 主要步骤：
   1) 将模型服务部署至目标环境；
   2) 根据工业互联网平台模型服务监测模块的反馈结果，基于预设的规则实现模型的迭代更新与持续部署。
c) 主要输出：模型 API、模型 SDK、指标分析结果、更新后的模型服务。

## 6 工业模型组件通用要求

## 6.1 概述

工业模型组件是指面向工业模型全生命周期，具有独立功能的软件功能模块，核心部分应包含模型开发组件、模型管理组件、模型服务组件和其他组件，见图2。

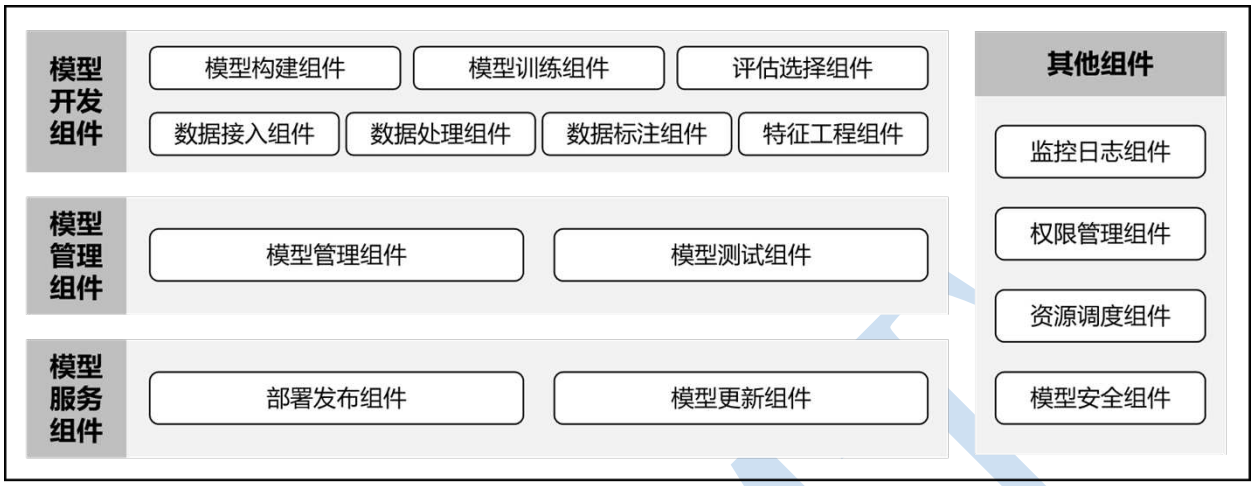



图2 工业模型组件总体架构图

## 6.2 工业模型开发阶段组件

### 6.2.1 数据接入组件

数据接入组件主要用于接入源数据并集中统一管理，其能力应包括下列内容。

a) 数据预览：组件应具备友好的用户界面，支持可视化查看、预览数据及标注信息。
b) 数据类型：组件应支持结构化、非结构化和半结构化等多类数据接入。
c) 数据来源：组件应支持从关系型数据库、非关系型数据库、文件系统等多类存储系统获取数据。
d) 接入方式：组件应支持以本地导入、外部链接等多种方式接入数据。
e) 模块集成：组件应支持与工业互联网平台工业大数据管理模块相关组件进行交互。

### 6.2.2 数据处理组件

数据处理组件主要用于对源数据进行检测和处理，纠正或删除已损坏、不准确或不适用的记录，其能力应包括下列内容。

a) 多类型：组件应支持对结构化数据、文本、图像、音频、视频等多类源数据的处理操作。
    1) 结构化数据：支持对结构化数据进行去重、无效值处理、缺失值处理等操作；
    2) 文本数据：支持对文本数据进行降低字频、生僻字添加等操作；
    3) 图像数据：支持对图像数据进行对比度/亮度调节、旋转、裁剪等操作；
    4) 视频数据：支持对视频数据进行抽帧等操作；
    5) 音频数据：支持对音频数据进行降噪等操作。
b) 数据处理：组件应具备对数据进行格式转换、去噪、去冗、修正、补齐（结构化数据），降低字频、生僻字添加（文本），亮度调节、对比度调节、旋转、裁剪（图像）、降噪（音频）、抽帧（视频）等数据综合处理能力。
c) 大批量：组件应具备大批量数据处理能力，能够承载一定容量规模的数据。
d) 数据集管理：组件应支持将处理完成的数据集发布为可用的公开/私有数据集，并支持进行数据集版本管理。

### 6.2.3 数据标注组件

数据标注组件主要用于对文本、图像等非结构化数据进行标注，其能力应包括下列内容。

a) 标注方式：组件应支持进行在线标注、智能标注、协作标注等多种方式，并提供热键支持等服务提升标注效率。

b) 标注类型：组件应支持对图片、文本、音频、视频、表格等多类数据进行标注工作。
c) 标签编辑：组件应支持进行标签/标签组的创建、修改、管理、删除等操作。
d) 标注准确：组件应具备良好的标注准确性，确保标注导出结果与与实际标注结果一致。

#### 6.2.4 特征工程组件

特征工程组件用于从源数据提取、构造和选择特征（或变量），其能力应包括下列内容。
a) 低代码：组件应支持以低代码方式创建特征工程项目。
b) 组件库：组件应具备丰富灵活的特征工程组件库，以满足不同的数据类型和业务场景的特征工程处理需求。
c) 可扩展：组件应支持用户扩展或自定义新的特征工程组件。
d) 可解释：组件宜提供特征可解释性工具，以图、表等形式可视化呈现特征影响和贡献度。

#### 6.2.5 模型构建组件

模型构建组件主要用于创建、编辑和管理模型开发任务，完成模型开发，其能力应包括下列内容。
a) 开发工具：组件应支持多语言开发环境并和可视化开发工具，包括 Python、C、R 等开发语言。
b) 开发环境：组件应集成主流的工业模型开发环境。
c) 开发方式：组件应支持代码式、可视化、离线上传等多种开发方式。
d) 算子组件：组件应支持基础算法的集成与扩展，宜包括机器学习、深度学习、数理统计、物理化学公式、信号处理等常用算子，以算法模板、可视化组件等形式呈现。
e) 增量训练：组件应支持用户选择已完成的训练任务为基准任务开展增量训练。
f) 可扩展：组件应提供标准化接口支持用户对开发环境、算法模板和可视化开发算子进行编辑和扩展。
g) 预置模型调参：组件宜支持用户选择合适的预训练模型、网络并通过参数配置完成模型开发，对于可配置参数的作用效果应给出简要说明。

#### 6.2.6 模型训练组件

模型训练组件主要用于创建、编辑和管理模型训练任务，完成模型训练，应包括下列内容。
a) 任务管理：组件应支持创建、编辑和管理模型训练任务。
b) 分布式训练：组件应支持分布式计算架构，加速大规模数据集和复杂模型的训练过程。
c) 资源调用：组件应支持调用计算资源以满足模型训练任务需要。

#### 6.2.7 评估选择组件

模型评估组件主要用于对训练完成的模型进行功能性测试与质量评价，应包括下列内容。
a) 评估指标：组件应支持准确率、精确率、召回率、ROC、AUC、PRC 等多种评估指标。
b) 评估范围：组件应支持对分类、回归、聚类、物体检测等多类模型的评估。
c) 评估方法：组件应支持留出法、自助法、交叉验证等模型评估策略。

### 6.3 工业模型管理阶段组件

#### 6.3.1 模型管理组件

模型管理组件主要用于通过标准化的模型接口，对自建、第三方模型进行集中纳管，应包括下列内容。
a) 模型来源：组件应支持本平台自训练和第三方导入的模型进行统一注册、纳管。
b) 模型格式：组件应支持第三方以模型文件、镜像等多种方式导入模型。
c) 版本管理：组件应支持对单个模型进行版本管理，管理内容宜包括版本的基础信息、环境依赖等内容并提供版本对比等服务。
d) 模型资产：组件应支持对模型文件、标签/参数等配置文件、模型操作/说明文档、模型评估指标报告等模型资产文件进行统一管理。
e) 部署包管理：组件应支持统一管理面向部署的模型镜像或 SDK 文件，支持基于模型文件、镜像等方式构建部署包。

f） 容器化管理：组件宜采用容器化的模型管理方式提升模型管理的效率。

### 6.3.2 模型测试组件

模型测试组件主要用于开展模型的非功能性测试，以评估模型是否具备在生产环境运行工作的能力。模型测试应支持进行模型结构测试、参数测试、集成测试、系统测试、业务测试、生产验证等非功能性测试。

## 6.4 工业模型服务阶段组件

### 6.4.1 部署发布组件

模型部署组件主要用于将训练好的工业模型部署至生产环境中，以API、SDK等方式向用户提供模型服务，应包括下列内容。

a） 部署方式：组件应支持以服务化部署、SDK 部署、模型文件下发等多种部署方式，并通过工业互联网平台的模型服务市场统一对外提供模型服务。
b） 推理框架：组件应支持主流模型推理框架，以满足服务化部署的模型推理任务需要。
c） 云边协同：支持以云边协同方式向边缘侧部署模型。
d） 可扩展：组件应支持用户扩展或自定义新的推理服务部署框架。
e） 容器化部署：组件宜支持容器化的模型部署方式提升模型部署的效率。

### 6.4.2 模型更新组件

模型更新组件用于根据模型服务运营监控结果自动触发并执行模型的迭代优化，并以用户无感知的方式完成模型更新，应包括下列内容。

a） 条件配置：组件应支持用户自定义配置模型更新的监控指标、阈值和触发条件。
b） 更新方式：组件应支持在线重训和离线重训等方式完成模型的更新。
c） 版本管理：组件应支持用户对更新前/后的模型进行版本对比、版本回滚等操作。

## 6.5 其他组件

### 6.5.1 监控日志组件

监控日志组件主要用于对基础设施、任务进程进行监控，并对工业模型相关组件的状态信息进行收集和展示，应包括下列内容。

a） 基础设施监控：组件应支持对主机或容器的健康状况、计算资源（CPU、GPU 等）使用率、I/O 占用率等进行监控和展示，保障运行环境的稳定。
b） 任务监控：组件应支持对模型训练过程中的训练时长、计算资源（CPU、GPU 等）使用率等指标进行统计、记录并展示。

**注：** 本标准仅对模型开发过程的相关指标进行监控；模型对外服务质量的监控能力由T/CPUMT FFFF—XXXX进行规定。

### 6.5.2 权限管理组件

权限管理组件主要用于管理工业模型项目相关的用户权限和访问控制，应包括下列内容。

a） 数据管理：组件应支持配置数据集的读/写权限，防止未授权用户访问、篡改数据集。
b） 算子管理：组件应支持配置开发算子的读/写权限，防止未授权用户访问、篡改开发算子。
c） 模型管理：组件应支持配置已纳管模型的读/写权限，防止未授权用户访问、篡改工业模型。

### 6.5.3 资源调度组件

资源调度组件主要用于对工业模型项目依赖的算力和存储资源面向多租户提供管理和调度服务，应包括下列内容。

a） 配额管理：在多租户体系下，组件应支持向租户分配资源池并独立管理。
b） 资源调度：组件应支持根据模型需求和系统负载情况动态地分配和回收计算资源，以最大限度提升资源利用率和系统性能。

### 6.5.4 模型安全组件

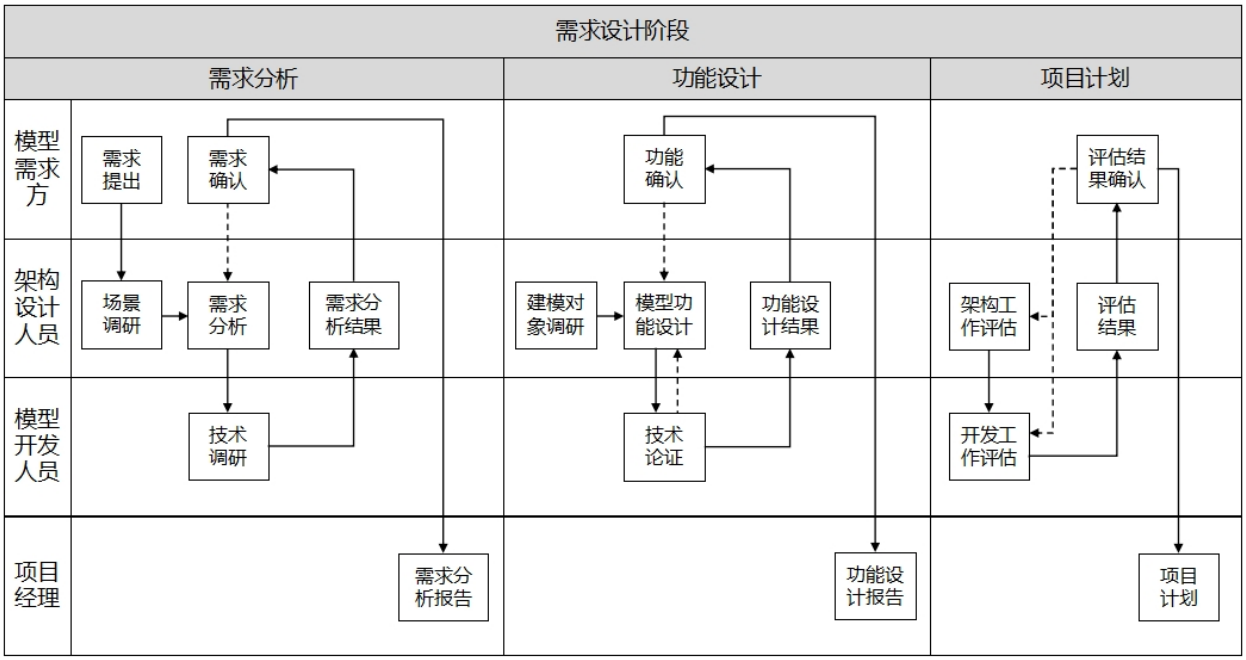
模型安全组件主要用于保障模型的机密性、完整性和可用性，应包括下列内容。

a) 访问控制：组件应支持身份验证和权限控制机制，防止未授权用户访问、篡改模型和相关数据。
b) 数据加密：组件应支持采用安全的加密算法，对传输和存储的数据进行加密，防止数据未授权访问。
c) 数据处理：组件应支持采用恶意样本识别等策略，防止数据投毒、数据篡改和恶意数据添加。
d) 模型开发：组件应支持采用模型签名、模型完整性检测等策略，防止模型篡改和逆向攻击。
e) 管理服务：组件应支持采用模型加密、模型水印等策略，保护已纳管和已部署模型的安全性和完整性。

# 附 录 A
（资料性）
工业模型需求设计阶段标准流程

## A.1 工业模型需求设计阶段标准流程

为更加高效、规范地开展工业模型需求设计阶段的各项工作，提供需求设计阶段标准流程，见图A.1：




注：“- - - →”表示当前节点异常状态处理过程。

图A.1 需求设计阶段标准流程

a) 需求分析：
   1) 模型需求方根据实际应用场景、业务需求，整理模型需求文件，包含对模型功能、性能需求的梳理，向架构设计人员提出需求；
   2) 架构设计人员依据模型需求文件，进行业务场景调研；
   3) 架构设计人员依据业务场景调研结果，进行用户需求分析；
   4) 模型开发人员根据需求分析结果，进行相关技术调研；
   5) 架构设计人员根据模整理形成需求分析结果；
   6) 模型需求方对架构设计人员反馈的需求分析结果进行确认。若结果满足实际业务需要，则由项目经理整理并形成需求分析报告；反之，则反馈至架构设计人员并重新进行需求分析工作。
b) 功能设计：
   1) 架构设计人员依据需求分析报告开展建模对象调研工作；
   2) 架构设计人员依据建模对象调研结果，开展模型功能设计工作，并将模型功能设计方案反馈至模型开发人员；
   3) 模型开发人员对于模型功能设计方案进行技术论证。若论证通过，则由架构设计人员整理功能设计结果；反之，则反馈至架构设计人员并重新进行模型功能设计工作；