国家互联网信息办公室｜工业和信息化部｜公安部｜国家广播电视总局

图解《人工智能生成合成内容标识办法》

《GB 45438-2025 网络安全技术 人工智能生成合成内容标识方法》

《网络安全标准实践指南-人工智能生成合成内容标识 服务提供者编码规则》

《生成式人工智能服务管理暂行办法》

技术规则与技术标准有机组合，提出人工智能标识管理的中国方案

《人工智能生成合成内容标识办法》

2025年3月7日，**国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局**联合发布《人工智能生成合成内容标识办法》（以下简称《标识办法》），自2025年9月1日起施行。

扩展解读

强制国标

《网络安全技术 人工智能生成合成内容标识方法》

该强制性国标由**中央网络安全和信息化委员会办公室**归口，委托**TC260全国网络安全标准化技术委员会**执行，于2025年9月1日实施。

实践指南

《网络安全标准实践指南-人工智能生成合成内容标识 服务提供者编码规则》

该指南由**全国网络安全标准化技术委员会秘书处**组织制定和发布，**指导其开展人工智能生成合成内容的文件元数据隐式标识工作**。

管理办法

《生成式人工智能服务管理暂行办法》

该办法经2023年5月23日**国家互联网信息办公室**2023年第12次室务会会议审议通过，并经**国家发展和改革委员会、教育部、科学技术部、工业和信息化部、公安部、国家广播电视总局**同意，自2023年8月15日起施行。

1. 图解四部门《人工智能生成合成内容标识办法》
2. 图解《GB 45438-2025 网络安全技术 人工智能生成合成内容标识方法》
3. 图解《网络安全标准实践指南——人工智能生成合成内容标识 服务提供者编码规则》
4. 图解网信办《生成式人工智能服务管理暂行办法》

《办法》将人工智能标识产业实践经验升维为有约束力的治理范式


中华人民共和国国家互联网信息办公室
Cyberspace Administration of China

首页 | 时政要闻 | 网信政务 | 互动服务 | 热点专题

关于印发《人工智能生成合成内容标识办法》的通知

关于印发《人工智能生成合成内容标识办法》的通知

国信办通字〔2025〕2号

各省、自治区、直辖市互联网信息办公室、通信管理局、公安厅（局）、广播电视局，新疆生产建设兵团互联网信息办公室、工业和信息化局、公安局、文化体育广电和旅游局：

为了促进人工智能健康发展，规范人工智能生成合成内容标识，保护公民、法人和其他组织合法权益，维护社会公共利益，国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局制定了《人工智能生成合成内容标识办法》，现印发给你们，请认真遵照执行。

国家互联网信息办公室

工业和信息化部

公安部

国家广播电视总局

2025年3月7日

2025年3月7日，国家互联网信息办公室、工业和信息化部、公安部、国家广播电视总局联合印发《人工智能生成合成内容标识办法》，自2025年9月1日起施行。

《办法》意义

《办法》旨在促进人工智能健康发展，规范人工智能生成合成内容标识，保护公民、法人和其他组织合法权益，维护社会公共利益。

网络安全法
2017年6月1日起施行

互联网信息服务深度合成管理规定
2023年1月10日施行

互联网信息服务算法推荐管理规定
2022年3月1日施行

生成式人工智能服务管理暂行办法
2023年8月15日施行

依据

《人工智能生成合成内容标识办法》

2025年9月1日起施行

《标识办法》出台背景

聚焦人工智能“生成合成内容标识”关键点

作为规范性文件，发挥内容标识提醒提示和监督溯源的技术作用，推动人工智能产业健康有序发展

1 人工智能安全问题引发社会关切

近年来，人工智能快速发展，为生成合成文本、图片、音频、视频等信息提供了便利工具，在促进经济社会发展的同时，**虚假信息、深度伪造、恶意输出等安全问题突出，加剧传统信息安全风险的动态性、渗透性和全局性。**联合国2024年《治理人工智能，助力造福人类》报告将“损害信息完整性”列为排名第一的风险。

2 生成合成内容标识已成国际惯例

- 无论欧盟《人工智能法》、澳大利亚《安全和负责任的人工智能咨询：澳大利亚政府的临时回应》，还是正在制定中的美国《编辑和深度伪造内容来源保护完整性法案》《数字内容溯源标识法案》、英国《人工智能（监管）法案》等提案，均对水印、标识义务有要求。
- 在我国《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》明确相关规定。尽管各国对数字水印、数字指纹、加密元数据等标识技术的具体要求不同，对平台标识义务的强度存在差异，**但在从事后内容审核转向生成端风险内嵌控制、提高人工智能治理能力的理念是一致的。**

3 《办法》以合理成本提高安全性

《办法》聚焦人工智能“生成合成内容标识”关键点，**通过标识提醒用户辨别虚假信息，明确相关服务主体的标识责任义务，规范内容制作、传播各环节标识行为，**以合理成本提高安全性，促进人工智能在文本对话、内容制作、辅助设计等各场景加快落地，**减轻人工智能生成合成技术滥用危害，防范利用人工智能技术制作传播虚假信息等风险行为，**推动人工智能健康有序发展。

《标识办法》制定的总体思路

细化已有规定

《互联网信息服务算法推荐管理规定》、《互联网信息服务深度合成管理规定》、《生成式人工智能服务管理暂行办法》中提出了标识有关要求，**《标识办法》作为规范性文件，进一步细化标识的具体实施规范。**

解决关键问题

《标识办法》**重点解决“哪些是生成的”“谁生成的”“从哪里生成的”等问题**，推动由生成到传播各环节的全流程安全管理，力争打造可信赖人工智能技术。

统筹发展和安全

考虑人工智能发展需要，**针对在文本中添加隐式标识，在多媒体文件中添加数字水印，仍是技术难点或可能增加企业成本，不作强制要求。**为降低平台企业标识成本，提升落地实施的可操作、可执行性，**创新提出文本符号标识、音频节奏标识、文件元数据标识等低成本实施**可行方法。

管理要求与技术标准一体化考虑

为推动《标识办法》落地实施，**强制性国家标准《网络安全技术 人工智能生成合成内容标识方法》同步发布，**更好地指导相关主体规范开展标识活动。

《办法》与强制性国标《网络安全技术 人工智能生成合成内容标识方法》的关系


ICS 35.030
CCS L 80

GB

中华人民共和国国家标准

GB 45438—2025

网络安全技术
人工智能生成合成内容标识方法

Cybersecurity technology—
Labeling method for content generated by artificial intelligence

2025-02-28发布　　2025-09-01实施

国家市场监督管理总局
国家标准化管理委员会　发布

《标识方法》是《办法》的配套强制性国标

二者是从技术规则到技术标准的有机组合，是兼具灵活性和约束力的制度创新

符合全球人工智能治理通用做法，**提出人工智能标识管理的中国方案**，为全球人工智能治理规则形成和兼容协调提供更多基础共识

立法层面

《标识办法》主要从立法层面提出管理要求，**明确生成合成内容制作传播各主体的责任义务**，为促进人工智能技术创新发展，对具体实施操作不做要求。

实施层面

《网络安全技术 人工智能生成合成内容标识方法》以强制性国家标准形式制定实施，**主要提出强制执行部分的标识具体实施方式和操作方法**，两者同步推出，于2025年9月1日同步实施，以更好地指导相关主体规范开展标识活动。

施行日期

《标识办法》与强制性国标施行日期均为2025年9月1日。考虑到企业需要时间充分理解相关规定和标准规范，坚持循序渐进的治理原则，针对性地开展能力建设和功能研发，基于标识技术实施的复杂程度、试点试行的实践经验，**设定《标识办法》和配套强制性国家标准6个月左右的施行过渡期**。

《标识办法》和《标识标准》的核心内容与治理逻辑

我国人工智能法治建设进程中的重要一步

体系性 技术性 协同性 融贯性 全流程 精细化

作为网络信息内容治理制度体系的有机组成部分，《标识办法》和《标识标准》是我国人工智能法治建设进程中的重要一步，既与《中华人民共和国网络安全法》一脉相承，又与《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》《生成式人工智能服务管理暂行办法》等有效衔接，构建的人工智能标识管理要求具有**体系性、技术性、协同性、融贯性、全流程、精细化等特点**，在明确治理标准和落地机制、促成治理实效方面，具备可操作性、落实基础和前景。

体系性、技术性：技治与法治互嵌运行，探索显隐双标识机制

技治与法治彼此的内嵌、牵连、塑造在人工智能标识管理中尤为明显。**《标识办法》援引的《互联网信息服务深度合成管理规定》涵盖人工智能输出的基本数据类型**，并明确各类信息内容的可标识性。**《标识标准》中显式标识与隐式标识的有机结合、综合运用**，使得通过技术检测机制识别人工智能生成合成内容成为可能，并在一定程度上具有确立主体行为性质的初步证据效力。

协同性、融贯性：多部门监管统筹协调，多制度衔接资源整合

在人工智能标识监管执法中，厘清相关主管部门管理职责，明确标识违规行为法律责任，强化标识管理与算法备案、安全评估等制度的衔接和程序协调至关重要。《标识办法》规定，服务提供者在履行算法备案、安全评估等手续时，应当按照本办法提供生成合成内容标识相关材料，并加强标识信息共享。**《标识办法》明确，违反本办法规定的，由网信、电信、公安、广播电视等有关主管部门依据职责，按照有关法律、行政法规、部门规章的规定予以处理。**

全流程、精细化：产业生态链治理，科学设定上下游主体义务

《标识办法》和《标识标准》从网络信息服务提供者这一整体视角切入，**进一步明晰适用标识义务的主体。**通过深入审视各主体在人工智能生态体系中的角色、定位、行为模式以及影响程度，精准区分为生成合成内容服务提供者、内容传播服务提供者这两大类主体。**同时，针对其他关键参与方，也进行明确规定，基本实现对生成合成内容上下游主体的全覆盖。**此外，《标识办法》体现出在数据内容创作与打击滥用间寻求平衡的考量，也表明在立法思路上对当前蓬勃发展的人工智能产业整体上的鼓励倾向。

《标识办法》和《标识标准》的制度意义

从制度构建角度审视

《标识办法》和《标识标准》的制度化进程，本质是将人工智能标识产业实践经验升维为有约束力的治理范式。

依托**“立法确认-标准转化-产业传导”**机制，我国构建起从技术共识迈向法律秩序的进阶之路。

这一过程契合我国人工智能立法秉持的总体思路，**即灵活运用既有规则，鼓励行业产业先行开展试点探索，同时针对痛点、难点问题精准发力，靶向施策，**其产生的影响也将辐射至全球人工智能治理的理念层面以及标准实践领域。

从规范协同角度审视

人工智能标识治理体系呈现“规则+技术”双轮驱动特征，覆盖从内容生产到传播的全链条治理。

《标识办法》**侧重构建“责任链闭环”**，通过主体资质审查、过程性义务等约束行为。《标识标准》**聚焦“技术锚点”建设**，针对关键技术要件确立具备可验证性的实施基准。我国《标准化法》规定，对保障人身健康和生命财产安全、国家安全、生态环境安全以及满足经济社会管理基本需要的技术要求，应当制定强制性国家标准。

将生成合成内容标识的最低要求上升为国家强制性标准，不仅是对《标识办法》的具体落地，更是表明在社会应用全过程中，人工智能标识及活动已与人身健康和生命财产安全、国家安全等紧密相连。

从技术治理效能角度审视

人工智能标识制度并非颠覆AI既有生产环境，而是致力于在现有生产基础上，提升人类对相关内容的辨识度以及机械处理的准确性。

《标识办法》和《标识标准》不仅及时响应将AI输出纳入监管视野的紧迫需求，标准化且有所区分的显式、隐式技术要求也为权利记载、算法备案、数据溯源、安全评估等工作，提供普适性、可量化度量与评价的操作方法及实用工具，**建立“生成即标识”的技术规范推动产业内生合规**，实现安全和发展间动态平衡。

这一制度体系也切实回应在人机互动、人机融合的新一轮技术跨越式发展中，**人类对真实与虚拟进行辨**别，从而形成行为、决策依据的基本需求。

炼石整理：《人工智能生成合成内容标识办法》总结构

炼石
CipherGateway

《人工智能生成合成内容标识办法》

第一部分 总则

1. 目的依据
2. 适用范围
3. 重要定义

第二部分 标识形式

4. 显式标识
5. 隐式标识

第三部分 责任义务

6. 规范传播活动的措施
7. 平台服务核验要求
8. 规范服务协议内容
9. 日志留存
10. 主动声明及违法行为
11. 他法遵循
12. 标识信息共享
13. 监管机构职责

第四部分 附则

14. 施行日期

衔接上位法，明确适用范围及重要定义

适用范围

符合《互联网信息服务算法推荐管理规定》、《互联网信息服务深度合成管理规定》、《生成式人工智能服务管理暂行办法》规定情形的网络信息服务提供者（以下简称“服务提供者”）开展人工智能生成合成内容标识活动，适用本办法。

解读：

《互联网信息服务深度合成管理规定》第二条明确“法律、行政法规另有规定的，依照其规定”，《生成式人工智能服务管理暂行办法》第二条第二款明确“国家对利用生成式人工智能服务从事新闻出版、影视制作、文艺创作等活动另有规定的，从其规定”。《标识办法》同样遵照上述适用条款，开展特定活动对内容标识另有规定的，从其规定。

重要定义

人工智能生成合成内容：指利用人工智能技术生成、合成的文本、图片、音频、视频、虚拟场景等信息。

人工智能生成合成内容标识包括显式标识和隐式标识。

显式标识：指在生成合成内容或者交互场景界面中添加的，以**文字、声音、图形等方式呈现并可以被用户明显感知到**的标识。

隐式标识：指**采取技术措施在生成合成内容文件数据中添加的，不易被用户明显感知**到的标识。

对文本、音频、图片、视频、虚拟场景等生成合成内容添加显式标识

显式标识

服务提供者提供的生成合成服务属于《互联网信息服务深度合成管理规定》**第十七条第一款情形的**，应当按照下列要求对生成合成内容添加显式标识：

《互联网信息服务深度合成管理规定》
第十七条（一）智能对话、智能写作等模拟自然人进行文本的生成或者编辑服务；

在文本的起始、末尾或者中间适当位置添加文字提示或者通用符号提示等标识，或者**在交互场景界面、文字周边**添加显著的提示标识；

在音频的起始、末尾或中间适当位置添加语音提示或者音频节奏提示等标识，或者**在交互场景界面中**添加显著的提示标识；

在图片的适当位置添加显著的提示标识；

在视频起始画面和视频播放周边的适当位置添加显著的提示标识，可以**在视频末尾和中间**适当位置添加显著的提示标识；

呈现虚拟场景时，**在起始画面的适当位置**添加显著的提示标识，可以**在虚拟场景持续服务过程中的适当位置**添加显著的提示标识；

其他生成合成服务场景根据自身应用特点添加显著的提示标识。

服务提供者提供生成合成内容下载、复制、导出等功能时，应当确保文件中含有满足要求的显式标识。

在生成合成内容的文件元数据中添加隐式标识

炼石 CipherGateway

隐式标识

服务提供者应当按照**《互联网信息服务深度合成管理规定》第十六条**的规定，在生成合成内容的文件元数据中添加隐式标识。

隐式标识包含生成合成内容属性信息、服务提供者名称或者编码、内容编号等制作要素信息。

《互联网信息服务深度合成管理规定》

第十六条 深度合成服务提供者对使用其服务生成或者编辑的信息内容，应当采取技术措施添加不影响用户使用的标识，并依照法律、行政法规和国家有关规定保存日志信息。

- **生成合成内容属性信息**
- **服务提供者名称**
- **编码**
- **内容编号**

鼓励服务提供者在生成合成内容中**添加数字水印等形式的隐式标识**。

文件元数据是指**按照特定编码格式嵌入到文件头部的描述性信息**，用于记录文件来源、属性、用途等信息内容。

规范内容传播环节的标识行为

提供网络信息内容传播服务的服务提供者应当采取下列措施，规范生成合成内容传播活动：

文件元数据标明为生成合成内容

核验文件元数据中是否含有隐式标识，**文件元数据明确标明为生成合成内容的**，采取适当方式在发布内容周边添加显著的提示标识，明确提醒公众该内容属于生成合成内容；

用户声明为生成合成内容

文件元数据中未核验到隐式标识，但用户声明为生成合成内容的，采取适当方式**在发布内容周边添加显著的提示标识**，提醒公众该内容可能为生成合成内容；

疑似生成合成内容

文件元数据中未核验到隐式标识，用户也未声明为生成合成内容，但提供网络信息内容传播服务的服务提供者检测到显式标识或者其他生成合成痕迹的，**识别为疑似生成合成内容，**采取适当方式在发布内容周边添加显著的提示标识，提醒公众该内容疑似生成合成内容；

提醒用户主动声明

提供必要的标识功能，并提醒用户主动声明发布内容中是否包含生成合成内容。

有前款第一项至第三项情形的，**应当在文件元数据中添加生成合成内容属性信息、传播平台名称或者编码、内容编号**等传播要素信息。

以行为主体为抓手进行义务分配，落实法律追责机制，实现上下游主体全覆盖

炼石 SecGateway

1.总则

2.标识形式

3.责任义务

4.附则

主体	情形	要求
任何组织和个人	三不得	• **不得**-恶意删除、篡改、伪造、隐匿本办法规定的生成合成内容标识 • **不得**-为他人实施上述恶意行为提供工具或者服务 • **不得**-通过不正当标识手段损害他人合法权益
互联网应用程序分发平台	在应用程序上架或上线审核时	应当要求互联网应用程序服务提供者说明是否提供人工智能生成合成服务。
	互联网应用程序服务提供者提供人工智能生成合成服务的	互联网应用程序分发平台应当核验其生成合成内容标识相关材料。
服务提供者	服务协议内容	应当在用户服务协议中明确说明生成合成内容标识的方法、样式等规范内容，并提示用户仔细阅读并理解相关的标识管理要求。
	用户申请服务提供者提供没有添加显式标识的生成合成内容的	服务提供者可以在通过用户协议明确用户的标识义务和使用责任后，提供不含显式标识的生成合成内容，并依法留存提供对象信息等相关**日志不少于六个月**。
	开展标识活动的	还应当符合相关法律、行政法规、部门规章和强制性国家标准的要求。
	在履行算法备案、安全评估等手续时	应当按照本办法提供生成合成内容标识相关材料，并加强标识信息共享，为防范打击相关违法犯罪活动提供支持和帮助。
用户	用户使用网络信息内容传播服务发布生成合成内容的	应当主动声明并使用服务提供者提供的标识功能进行标识。

多部门监管统筹协调，明确施行日期

监管主体责任

- 违反本办法规定的，由**网信、电信、公安和广播电视**等有关主管部门依据职责，按照有关法律、行政法规、部门规章的规定予以处理。

施行日期

- 本办法自2025年9月1日起施行。

1. 图解四部门《人工智能生成合成内容标识办法》
2. 图解《GB 45438-2025 网络安全技术 人工智能生成合成内容标识方法》
3. 图解《网络安全标准实践指南——人工智能生成合成内容标识 服务提供者编码规则》
4. 图解网信办《生成式人工智能服务管理暂行办法》

GB 45438-2025是配套《人工智能生成合成内容标识办法》的强制性国标

本标准是配套《人工智能生成合成内容标识办法》的强制性国家标准，对防范人工智能生成合成内容引发安全风险、提升人工智能安全水平起到规范作用，促进人工智能行业安全发展。

适用范围

- 本文件规定了人工智能生成合成内容标识方法。
- 本文件**适用于生成合成服务提供者和内容传播服务提供者开展人工智能生成合成内容标识活动。**

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。

- **GB18030—2022 信息技术 中文编码字符集**

明确八大术语定义

炼石 CipherGateway

人工智能生成合成内容
contentgeneratedbyartificialintelligence

利用人工智能技术生成、合成的文本、图片、音频、视频、虚拟场景等信息。

人工智能生成合成内容标识
labelofcontentgeneratedbyartificialintelligence

表明某个内容是人工智能生成合成内容(3.1)的信息。
注：标识可包含生成合成服务提供者和内容传播服务提供者等信息。

显式标识
explicitlabel

在人工智能生成合成内容或交互场景界面中添加的，以文字、声音、图形等方式呈现并可被用户明显感知到的标识。

隐式标识
implicitlabel

采取技术措施在人工智能生成合成内容文件数据中添加的，不易被用户明显感知到的标识。

文件元数据
filemetadata

按照特定编码格式嵌入到文件中的描述性数据，用于记录文件来源、属性、用途、版权等信息。

文件元数据隐式标识
filemetadataimplicitlabel

在人工智能生成合成内容文件元数据中添加的隐式标识。

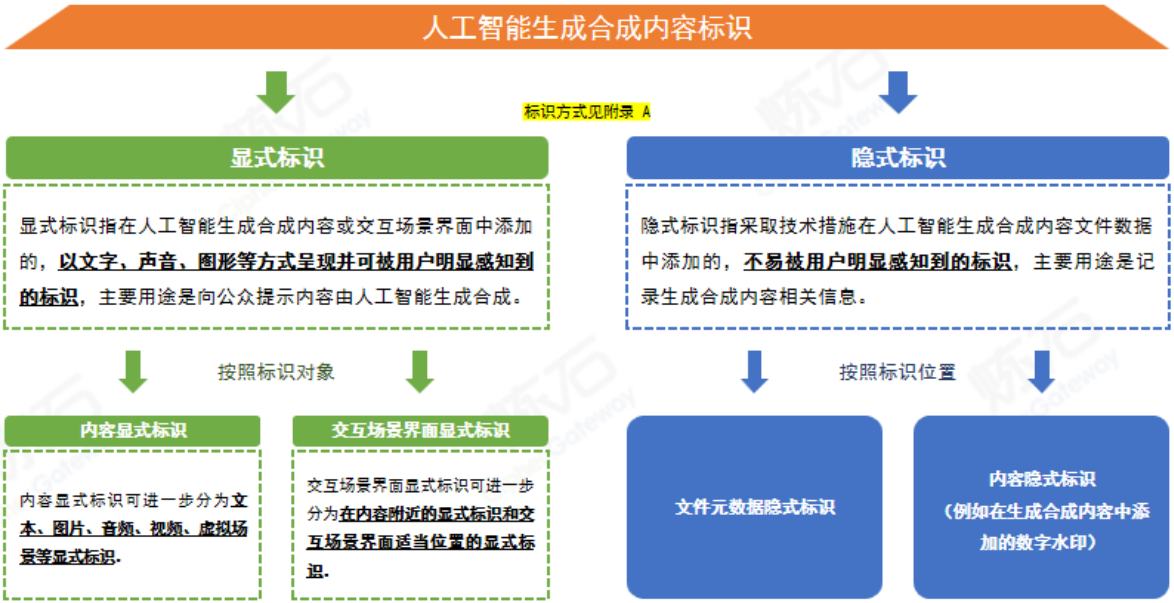
人工智能生成合成内容服务提供者
artificialintelligencecontentgenerationserviceprovider

利用人工智能技术(包括通过提供可编程接口等方式)向公众提供生成合成文本、图片、音频、视频、虚拟场景等服务的组织或个人。

网络信息内容传播服务提供者
internetinformationcontentpropagationserviceprovide

提供网络信息内容传播服务的网络信息服务提供者。

人工智能生成合成内容标识包含显式标识和隐式标识二种方式



附录A：标识方式

显式标识

显式标识包括内容显式标识、交互场景界面显式标识等，如图 A.1所示。

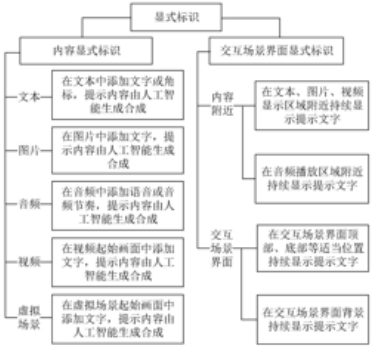


图 A.1　显式标识

隐式标识

隐式标识包括文件元数据隐式标识、内容隐式标识等，如图 A.2所示。

内容隐式标识是在人工智能生成合成内容数据中添加的数字水印等形式的隐式标识，在图 A.2中用虚线框表示，在本文件中不做要求。

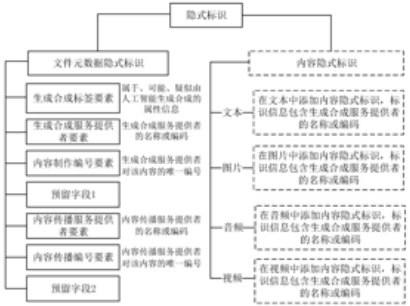


图 A.2　隐式标识

附录B：显式标识典型应用场景

显式标识典型应用场景包括在以下服务中，可能导致公众混淆或者误认的：

a. 智能对话、智能写作等模拟自然人进行文本的生成或者编辑服务；

b. 合成人声、仿声等语音生成或者显著改变个人身份特征的编辑服务；

c. 人脸生成、人脸替换等人物图像、视频生成或者显著改变个人身份特征的编辑服务；

d. 人脸操控、姿态操控等人物图像、视频生成或者显著改变个人身份特征的编辑服务；

e. 沉浸式拟真场景等生成或者编辑服务；

f. 文生图片等图片内容生成服务；

g. 音乐创作等音频内容生成服务；

h. 文生视频、图生视频等视频内容生成服务；

i. 其他具有生成或者显著改变信息内容功能的服务。

文本、图片内容显式标识

① 文本内容显式标识

文本内容显式标识方法如下。

a）文本内容显式标识**应采用文字或角标形式**。

b）文字形式的文本内容显式标识应同时包含以下要素：

1）人工智能要素：包含“人工智能”或“AI”，表明使用人工智能技术；

注：“AI”是人工智能(ArtificialIntelligence)的缩写。

2）生成合成要素：包含“生成”和/或“合成”，表明内容制作方式为生成和/或合成。

c）角标形式的文本内容显式标识**应包含“AI”**。

d）文本内容显式标识**应位于以下一个或多个位置：**

1）文本的起始位置；

2）文本的末尾位置；

3）文本的中间适当位置。

e）文本内容显式标识**使用的字型和颜色应清晰可辨**。

文本内容显式标识示例见附录 C的 C.1。

② 图片内容显式标识

图片内容显式标识方法如下。

a）图片内容显式标识**应采用文字提示**。

b）图片内容显式标识应同时包含以下要素：

1）人工智能要素：包含“人工智能”或“AI”，表明使用人工智能技术；

2）生成合成要素：包含“生成”和/或“合成”，表明内容制作方式为生成和/或合成。

c）图片内容显式标识**应位于图片的边或角**。

d）图片内容显式标识**使用的字型应清晰可辨**。

e）图片内容显式标识的**文字高度不应低于画面最短边长度的5%**。

注：非矩形图片的最短边指可完全将该图片包含在内部的矩形的最短边。

图片内容显式标识示例见 C.2。

音频、视频内容显式标识

③ 音频内容显式标识

音频内容显式标识方法如下。

a）音频内容显式标识**应采用语音标识或音频节奏标识**。

b）语音标识应包含以下要素：

1）人工智能要素：包含“人工智能”或“AI”，表明使用人工智能技术；

2）生成合成要素：包含“生成”和/或“合成”，表明内容制作方式为生成和/或合成。

c）音频节奏标识**应为“短长　短短”的节奏**。

注1：“短长　短短”节奏为“AI”的摩斯码表示。

d）音频内容显式标识**应位于以下一个或多个位置**：

1）音频的起始位置；

2）音频的末尾位置；

3）音频的中间适当位置。

注2：音频的起始位置位于生成合成的音频内容开始之前。音频的末尾位置位于生成合成的音频内容结束之后。

注 3：在智能语音助手、智能客服、智能导航等音频的高频交互场景中，音频的起始位置、末尾位置是指一轮交互的起始位置和末尾位置。

e）语音标识**应使用正常语速**。

注4：汉语正常语速约在 120字/min~160 字/min。

f）节奏标识**应清晰可辨**。

音频内容显式标识示例见C.3。

④ 视频内容显式标识

视频内容显式标识方法如下。

a）视频内容显式标识**应采用文字提示**。

b）视频内容显式标识应同时包含以下要素：

1）人工智能要素：包含“人工智能”或“AI”，表明使用人工智能技术；

2）生成合成要素：包含“生成”和/或“合成”，表明内容制作方式为生成和/或合成。

c）视频内容显式标识**应位于视频起始画面，可位于视频末尾和中间适当位置。**

d）视频内容显式标识**应位于视频画面的边或角**。

e）视频内容显式标识使用的字型**应清晰可辨**。

f）视频内容显式标识的**文字高度不应低于画面最短边长度的5%。**

g）在视频正常播放速度下，**视频内容显式标识持续时间不应少于2s。**

视频内容显式标识示例见 C.4。

虚拟场景、交互场景界面显式标识

5 虚拟场景显式标识

虚拟场景显式标识方法如下。

a）虚拟场景显式标识**应采用文字提示**。

b）虚拟场景显式标识应同时包含以下要素：

1）人工智能要素：包含“人工智能”或“AI”，表明使用人工智能技术；

2）生成合成要素：包含“生成”和/或“合成”，表明内容制作方式为生成和/或合成。

c）虚拟场景显式标识**应位于虚拟场景起始画面，可位于虚拟场景持续服务过程中的适当位置。**

d）位于虚拟场景起始画面的虚拟场景显式标识**应位于画面的边或角。**

e）虚拟场景显式标识使用的字型**应清晰可辨**。

f）虚拟场景显式标识的**文字高度不应低于画面最短边长度的5%。**

6 交互场景界面显式标识

交互场景界面显式标识方法如下。

a）交互场景界面显式标识**应采用文字提示**。

b）交互场景界面显式标识应同时包含以下要素：

1）人工智能要素：包含“人工智能”或“AI”，表明使用人工智能技术；

2）生成合成要素：包含“生成”和/或“合成”，表明内容制作方式为生成和/或合成。

c）交互场景界面显式标识**应采取以下一种或多种方式**：

1）在内容附近持续显示提示文字；

2）在交互场景界面顶部、底部、背景等适当位置持续显示提示文字。

d）交互场景界面显式标识使用的字型和颜色**应清晰可辨**。

交互场景界面显式标识示例见附录 D。

文件元数据、内容隐式标识

1 文件元数据隐式标识

文件元数据隐式标识方法如下。

a）隐式标识应包括以下要素：

1）生成合成标签要素：内容的人工智能生成合成属性信息；

2）生成合成服务提供者要素：生成合成服务提供者的名称或编码；

3）内容制作编号要素：生成合成服务提供者对该内容的唯一编号；

4）内容传播服务提供者要素：内容传播服务提供者的名称或编码；

5）内容传播编号要素：内容传播服务提供者对该内容的唯一编号。

b）文件元数据隐式标识格式**应符合附录 E**。

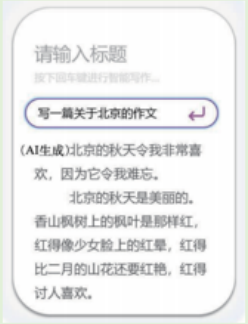
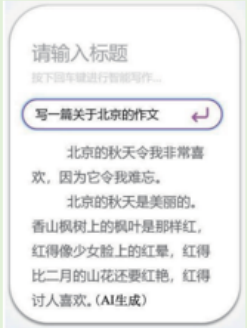
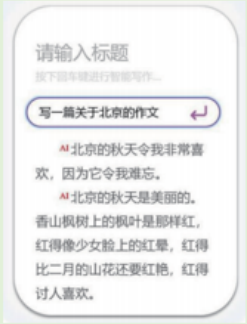
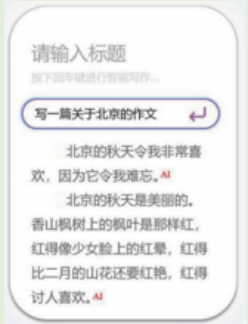
c）人工智能生成合成的内容文件中，**应仅保留一份文件元数据隐式标识。**

文件元数据隐式标识示例见附录 F。

2 内容隐式标识

内容隐式标识**准许采用数字水印等形式**。

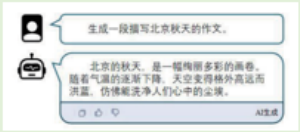
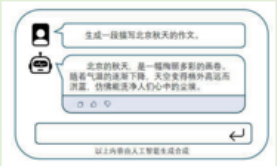
附录C：内容显式标识示例 – 文本内容显式标识

类别	文本内容显式标识			
形式	文字形式		角标形式	
位置	使用“AI生成”提示文字，位于**文本起始位置**的内容显式标识示例见图 C.1。	使用“AI生成”提示文字，位于**文本末尾位置**的内容显式标识示例见图 C.2。	使用“AI”提示角标，位于**文本起始位置**的内容显式标识示例见图 C.3。	使用“AI”提示角标，位于**文本末尾位置**的内容显式标识示例见图 C.4。
示例图				
	图 C.1 位于文本起始位置的文字形式内容显式标识示例	图 C.2 位于文本末尾位置的文字形式内容显式标识示例	图 C.3 位于文本起始位置的角标形式内容显式标识示例	图 C.4 位于文本末尾位置的角标形式内容显式标识示例

附录C：内容显式标识示例 – 图片内容、音频内容、视频内容显式标识

类别	图片内容显式标识	音频内容显式标识	视频内容显式标识
位置	使用“人工智能生成合成”提示文字，位于图片右下角的内容显式标识示例见图 C.5。	在音频起始位置添加语音或音频节奏标识示例见图 C.6。	使用“人工智能生成合成”提示文字，位于视频起始画面右下角的内容显式标识示例见图 C.7。
示例图			
	图 C.5 位于图片右下角的内容显式标识示例	图 C.6 位于音频起始位置的语音或音频节奏标识示例	图 C.7 位于视频起始画面右下角的内容显式标识示例

附录D：交互场景界面显式标识示例

类别	在内容附近持续显式		在交互场景界面适当位置持续显示	
形式	内容显示区域附近	音频播放区域附近	交互场景界面底部	交互场景界面背景
位置	使用包含“AI生成”提示文字，位于文本内容显示区域附近的交互场景界面显式标识示例见图D.1。	使用包含“AI生成”提示文字，位于音频播放区域附近的交互场景界面显式标识示例见图D.2。	使用包含“人工智能生成合成”提示文字，位于交互场景界面底部的交互场景界面显式标识示例见图 D.3。	使用包含“人工智能生成合成”提示文字，位于交互场景界面背景的交互场景界面显式标识示例见图 D.4。
示例图				
	图 D.1 位于文本内容显示区域附近的交互场景界面显式标识示例	图 D.2 位于音频播放区域附近的交互场景界面显式标识示例	图 D.3 位于交互场景界面底部的交互场景界面显式标识示例	图 图 D.4 位于交互场景界面背景的交互场景界面显式标识示例

附录E：文件元数据隐式标识格式[1]

文件元数据隐式标识格式要求如下

a）在文件元数据中添加隐式标识扩展字段，字段名称或关键词中应包含“AIGC”。

b）隐式标识扩展字段的值，应为符合以下格式的字符串。

```
{"AIGC": {"Label": "value1", "ContentProducer": "value2", "ProduceID": "
value3", "ReservedCode1": "value4", "ContentPropagator": "value5", "PropagateID":
"value6", "ReservedCode2": "value7"}}
```

注1：生成合成服务提供者对人工智能生成合成内容首次写入文件元数据隐式标识时，内容传播服务提供者要素与生成合成服务提供者要素一致，内容传播编号要素与内容制作编号要素一致。

c）生成合成标签要素由 Label表示，取值为value1，应符合以下要求。

1）存储内容属于、可能、疑似为人工智能生成合成的属性信息：属于人工智能生成合成内容的，value1 的值取1；可能为人工智能生成合成内容的，value1 的值取2；疑似为人工智能生成合成内容的，value1 的值取3。

2）类型为字符串。

d）生成合成服务提供者要素由 ContentProducer表示，取值为value2，应符合以下要求：

1）存储生成合成服务提供者的名称或编码；

2）类型为字符串。

e）内容制作编号要素由 ProduceID表示，取值为value3，应符合以下要求：

1）存储生成合成服务提供者对该内容的唯一编号；

2）类型为字符串。

附录E：文件元数据隐式标识格式[2]

文件元数据隐式标识格式要求如下

f）预留字段1由 ReservedCode1表示，取值为value4，要求如下：

1）可存储用于生成合成服务提供者自主开展安全防护，保护内容、标识完整性的信息；

2）类型应为字符串。

注2：生成合成服务提供者使用预留字段1进行文件元数据隐式标识安全防护的示例见附录 F的 F.4。

g）内容传播服务提供者要素由 ContentPropagator表示，取值为value5，应符合以下要求：

1）存储内容传播服务提供者的名称或编码；

2）类型为字符串。

h）内容传播编号要素由 PropagateID表示，取值为value6，应符合以下要求：

1）存储内容传播服务提供者对该内容的唯一编号；

2）类型为字符串。

i）预留字段2由 ReservedCode2表示，取值为value7，要求如下：

1）可存储用于内容传播服务提供者自主开展安全防护，保护内容、标识完整性的信息；

2）类型应为字符串。

j）c)~i)中各要素的值应主要由 GB18030—2022中码位为0x21、0x23~0x5B、0x5D~0x7E 的字符以及\"构成。

注3：码位0x21、0x23~0x5B、0x5D~0x7E包括除"和\外的所有单字节编码字符，\"为转义字符，用于表示"。

附录F：文件元数据隐式标识示例

类别	图片文件元数据隐式标识示例	音频文件元数据隐式标识示例	视频文件元数据隐式标识示例	使用数字签名的安全防护示例
位置	**属于**人工智能生成合成的**图片文件**元数据隐式标识示例见图 F.1的“AIGC”部分。	**可能为**人工智能生成合成的**音频文件**元数据隐式标识示例见图 F.2中的“AIGC”部分。	**疑似为**人工智能生成合成的**视频文件**元数据隐式标识示例见图 F.3中的“AIGC”部分。	**生成合成服务提供者使用杂凑算法对文件元数据信息进行数字签名，并将结果存储在预留字段1中**的示例如下所示。
示例图				"ReservedCode1"："e862483430d978cbf828b8b24296ef9328d843a0"
	图 F.1 图片文件元数据隐式标识示例	图 F.2 音频文件元数据隐式标识示例	图 F.3 视频文件元数据隐式标识示例	

1. 图解四部门《人工智能生成合成内容标识办法》
2. 图解《GB 45438-2025 网络安全技术 人工智能生成合成内容标识方法》
3. 图解《网络安全标准实践指南-人工智能生成合成内容标识 服务提供者编码规则》
4. 图解网信办《生成式人工智能服务管理暂行办法》

《实践指南》指导开展人工智能生成合成内容的文件元数据隐式标识工作


TC260-PG-20252A

网络安全标准实践指南

——人工智能生成合成内容标识 服务提供者编码规则

（V1.0-202503）

全国网络安全标准化技术委员会秘书处

2025年3月

本文档可从以下网址获得：

www.tc260.org.cn/

全国网络安全标准化技术委员会

《网络安全标准实践指南》（以下简称《实践指南》）是全国网络安全标准化技术委员会（以下简称“网安标委”）秘书处组织制定和发布的标准相关技术文件，**旨在围绕网络安全法律法规政策、标准、网络安全热点和事件等主题，宣传网络安全相关标准及知识，提供标准化实践指引。**

根据《人工智能生成合成内容标识办法》和强制性国家标准《网络安全技术，人工智能生成合成内容标识方法》的要求，**本指南为人工智能生成合成内容服务提供者和网络信息内容传播服务提供者提供编码规则，指导其开展人工智能生成合成内容的文件元数据隐式标识工作。**

《人工智能生成合成内容标识办法》 **《网络安全技术，人工智能生成合成内容标识方法》**

根据要求

《网络安全标准实践指南——人工智能生成合成内容标识 服务提供者编码规则》

人工智能生成合成内容标识服务提供者编码规则的适用范围

适用范围

- 本实践指南给出了人工智能生成合成内容服务提供者和网络信息内容传播服务提供者的**编码结构**和**赋码规则**。
- 本实践指南适用于指导人工智能生成合成内容服务提供者和网络信息内容传播服务提供者，**开展人工智能生成合成内容的文件元数据隐式标识活动**。

《实践指南》中涉及到的术语定义

服务提供者
service provider

生成合成服务提供者和/或内容传播服务提供者

人工智能生成合成内容服务提供者
artificial intelligencecontent generation service provider

生成合成服务提供者

利用人工智能技术（包括通过提供可编程接口等方式）向公众提供生成合成**文本**、**图片**、**音频**、**视频**、**虚拟场景**等内容服务的组织、个人。

网络信息内容传播服务提供者
internet information contentpropagation service provider

内容传播服务提供者

提供网络信息内容传播服务的**组织**、**个人**。

服务提供者编码的编码结构

服务提供者编码由二十七位的阿拉伯数字或大写英文字母组成，包括第 1 位～第 2 位**标识格式定义码**，第 3 位～第 22 位**主体标识码**，第 23 位～第 27 位**服务扩展码**三个部分。

表 1 编码构成

编码序号	1	2	3	4	…	21	22	23	24	25	26	27
编码	×	×	×	×	×	×	×	×	×	×	×	×
说明	标识格式定义码2位		主体标识码 20位					服务扩展码 5位				

服务提供者的编码规则

炼石 CipherGateway

赋码规则

编码序号	编码	编码规则
1	0	对于采用本指南编制的编码，标识格式定义码固定赋值为阿拉伯数字“00”。
2	0	
3	1	**主体类型分为组织和个人两类。**主体类型为组织赋值为“1”，主体类型为个人的赋值为“2”。
4	2	组织主体采用**统一社会信用代码**与其绑定时，赋值为阿拉伯数字“1”。 个人主体采用**公民身份号码**与其绑定时，赋值为阿拉伯数字“1”； 采用**移动通信终端号码**与其绑定时，赋值为阿拉伯数字“2”； 采用**护照号码**与其绑定时，赋值为阿拉伯数字“3”； 采用**网号**与其绑定时，赋值为阿拉伯数字“4”。
5	3	组织主体编码采用统一社会信用代码。 个人主体编码采用公民身份号码，或移动通信终端号码，或护照号码，或网号： a）采用**公民身份号码**时，主体编码为 18 位公民身份号码； b）采用**移动通信终端号码**编码时，主体编码为 11 位移动通信终端号码的组合和 7 位“0”的组合； c）采用**护照号码**编码时，主体编码为 9 位护照号码与 9 位“0”的组合； d）采用**网号**编码时，主体编码为 10 位网号与 8 位“0”的组合。
…	…	
22	4	
23	5	**服务提供者可不编写服务扩展码。**此时，服务扩展码填写阿拉伯数字“00000”。 第 23 位，记录服务类型。服务类型为生成合成服务，赋值为阿拉伯数字“1”；服务类型为内容传播服务，赋值为阿拉伯数字“2”。 第 24 位~第 27 位，记录模型/应用码，由服务提供者自行编写。
…	…	
27	6	

服务提供者的编码规则对应示例

赋码示例

0 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 1 2 3 4 5

指南编码	00

主体类型	赋值
组织	1
个人	2

	与主体绑定方式	赋值
组织主体绑定方式编码	统一社会信用代码	1
个人主体绑定方式编码	公民身份号码	1
	移动通信终端号码	2
	护照号码	3
	网号	4

	与主体绑定方式	赋值
组织主体编码	统一社会信用代码	91350100M000100Y43
个人主体编码	公民身份号码	110105194912310002X
	移动通信终端号码	12233334444000000
	护照号码	EF126089200000000000
	网号	123456789A00000000

自行编写

服务类型	赋值
生成合成服务	1
内容传播服务	2

附录 A 服务提供者编码示例

组织编码示例

组织统一社会信用代码	服务扩展码	服务提供者编码
91350100M000100Y43	不编写时	001191350100M000100Y4300000

个人编码示例

示例1	公民身份号码作为个人主体绑定方式	服务类型为生成合成服务	服务提供者编码
	1101051949123100 2X	0832	00211101051949123100 2X10832
示例2	移动通信终端号码作为个人主体绑定方式	服务类型为内容传播服务	服务提供者编码
	12233334444	0832	00221223333444400000002083 2
示例3	护照号码作为个人主体绑定方式	服务扩展码	服务提供者编码
	EF1260892	不编写时	0023EF1260892000000000000 0
示例4	个人使用网 号作为个人主体绑定方 式	服务扩展码	服务提供者编码
	123456789A	不编写时	0024123456789A000000000000 0

注： 附录 A 中使用的统一社会信用代码、公民身份号码、移动通信终端号码、 护照号码和网号为服务提供者编码示例，与相关编码的规则无关，且不指向特定主体。

1. 图解四部门《人工智能生成合成内容标识办法》
2. 图解《GB 45438-2025 网络安全技术 人工智能生成合成内容标识方法》
3. 图解《网络安全标准实践指南——人工智能生成合成内容标识 服务提供者编码规则》
4. 图解网信办《生成式人工智能服务管理暂行办法》

新法将促进生成式人工智能健康发展和规范应用


中华人民共和国国家互联网信息办公室
Cyberspace Administration of China

首页 | 时政要闻 | 网信政务 | 互动服务 | 热点专题

国家网信办等七部门联合公布《生成式人工智能服务管理暂行办法》

2023年07月13日 15:00 来源：中国网信网 【打印】【纠错】

近日，国家网信办联合国家发展改革委、教育部、科技部、工业和信息化部、公安部、广电总局公布《生成式人工智能服务管理暂行办法》（以下称《办法》），自2023年8月15日起施行。国家互联网信息办公室有关负责人表示，出台《办法》，旨在促进生成式人工智能健康发展和规范应用，维护国家安全和社会公共利益，保护公民、法人和其他组织的合法权益。

近年来，生成式人工智能技术快速发展，为经济社会发展带来新机遇的同时，也产生了传播虚假信息、侵害个人信息权益、数据安全和偏见歧视等问题，如何预判生成式人工智能发展和安全引起各方关注。出台《办法》，既是促进生成式人工智能健康发展的重要要求，也是防范生成式人工智能服务风险的现实需要。

《办法》提出国家坚持发展和安全并重、促进创新和依法治理相结合的原则，采取有效措施鼓励生成式人工智能创新发展，对生成式人工智能服务实行包容审慎和分类分级监管，明确了提供和使用生成式人工智能服务总体要求。提出了促进生成式人工智能技术发展的具体措施，明确了训练数据处理活动和数据标注等要求。规定了生成式人工智能服务规范，明确生成式人工智能服务提供者应当采取有效措施防范未成年人用户过度依赖或者沉迷生成式人工智能服务，按照《互联网信息服务深度合成管理规定》对图片、视频等生成内容进行标识。发现违法内容应当及时采取处置措施等。此外，还规定了安全评估、算法备案、投诉举报等制度，明确了法律责任。

国家互联网信息办公室有关负责人指出，生成式人工智能服务的发展与治理需要政府、企业、社会、网民等多方参与，共同促进生成式人工智能健康发展，让生成式人工智能技术更好地造福人民。

2023年7月，国家网信办联合国家发展改革委、教育部、科技部、工业和信息化部、公安部、广电总局公布《生成式人工智能服务管理暂行办法》（以下称《办法》），自2023年8月15日起施行。国家互联网信息办公室有关负责人表示，出台《办法》，旨在促进生成式人工智能健康发展和规范应用，维护国家安全和社会公共利益，保护公民、法人和其他组织的合法权益。

《生成式人工智能服务管理暂行办法》整体架构

炼石 CipherGateway

生成式人工智能服务管理暂行办法

一、总体原则

- 1.立法目的和依据
- 2. 适用范围、关键定义
- 3. 坚持创新和依法治理相结合原则

二、技术发展与治理

- 4.法律法规和公序良俗遵循
- 5.鼓励各行业领域创新应用
- 6.鼓励基础技术自主创新
- 7. 依法开展训练数据处理活动
- 8. 数据标注规定

三、服务规范

- 9.履行网络信息安全义务
- 10.未成年人防沉迷
- 11.依法保护个人信息
- 12.对生成内容进行标识
- 13. 提供安全、稳定、持续服务
- 14. 发现违法内容的应及时处理
- 15. 建立健全投诉、举报机制

四、监督检查和法律责任

- 16. 各部门职责
- 17. 按规定开展安全评估
- 18. 不合法合规有权投诉、举报
- 19. 提供者依法配合监督检查
- 20. 境外向境内提供服务需合法合规
- 21. 违规处罚

五、附　则

- 22.用语定义
- 23. 行政许可
- 24. 施行日期

网络监管日益成熟，实现技术发展与安全合规同步推动

划定适用范围、给出权威释义

1. **完善监管框架，**制定依据源于《网络安全法》《数据安全法》《个人信息保护法》《科学技术进步法》四大上位法，参照细化了《互联网信息服务算法推荐管理规定》《互联网信息服务深度合成管理规定》《互联网信息服务管理办法》等制度

2. **划定适用范围，明确关键定义，**鼓励行业开展基础技术自主创新、国际合作

明确权责划分、责任主体

1. **明确提供者的责任和义务**

2. **《办法》贯彻落实《个人信息保护法》，强化个人信息保护要求**
   - *提出不得侵害他人肖像权、名誉权、荣誉权、隐私权和个人信息权益*
   - *训练数据处理活动涉及个人信息的，应当取得个人同意或者符合法律、行政法规规定的其他情形*
   - *涉及个人信息的，依法承担个人信息处理者责任，履行个人信息保护义务。*
   - *参与生成式人工智能服务安全评估和监督检查的相关机构和人员对在履行职责中知悉的国家秘密、商业秘密、个人隐私和个人信息应当依法予以保密，不得泄露或者非法向他人提供。*

建立多项机制、丰富管理措施

- **开展安全评估：**《办法》要求提供具有舆论属性或者社会动员能力的生成式人工智能服务的，应当按照国家有关规定开展安全评估

- **丰富管理措施：**通过**履行网络信息安全义务、未成年人防沉迷、依法保护个人信息、对生成内容进行标识、建立健全投诉举报机制等措施和手段，增强生成式人工智能合法性、合规性和合理性**

第一条：《办法》制定的目的和依据是什么？

构建完善的生成式人工智能法制化监管体系，促进生成式人工智能技术健康发展和规范应用

上位法 《网络安全法》

上位法 《数据安全法》

上位法 《个人信息保护法》

上位法 《科学技术进步法》

参考依据

《互联网信息服务深度合成管理规定》
2023.1.10起施行

《互联网信息服务算法推荐管理规定》
2022.3.1起施行

《网络信息内容生态治理规定》
2020.3.1起施行

《具有舆论属性或社会动员能力的互联网信息服务安全评估规定》
2018.11.30起施行

《互联网信息服务管理办法》
2011.1.8修订

《生成式人工智能服务管理暂行办法》

第二、三条：《办法》适用范围是什么？体现什么样的基本原则？

什么样的组织和个人适用《办法》？

利用生成式人工智能技术向境内公众提供生成文本、图片、音频、视频等内容的服务，适用本办法。国家对利用生成式人工智能服务从事新闻出版、影视制作、文艺创作等活动另有规定的，从其规定。

什么是生成式人工智能技术？

生成式人工智能

是指具有文本、图片、音频、视频等内容生成能力的模型及相关技术

体现什么样的基本原则？

国家

↓

坚持发展和安全并重、促进创新和依法治理相结合的原则

↓

采取有效措施鼓励生成式人工智能创新发展，对生成式人工智能服务实行包容审慎和分类分级监管

第四条：提供和使用生成式人工智能服务应当遵守哪些要求？

炼石 CipherGateway

提供和使用生成式人工智能服务，应当遵守法律、行政法规，尊重社会公德和伦理道德，遵守以下规定：

1 社会主义核心价值观

- 坚持社会主义核心价值观，不得生成煽动颠覆国家政权、推翻社会主义制度，危害国家安全和利益、损害国家形象，煽动分裂国家、破坏国家统一和社会稳定，宣扬恐怖主义、极端主义，宣扬民族仇恨、民族歧视，暴力、淫秽色情，以及虚假有害信息等法律、行政法规禁止的内容；

2 防止歧视

- 在算法设计、训练数据选择、模型生成和优化、提供服务等过程中，采取有效措施防止产生民族、信仰、国别、地域、性别、年龄、职业、健康等歧视

3 不得实施不公平竞争

- 尊重知识产权、商业道德，保守商业秘密，不得利用算法、数据、平台等优势，实施垄断和不正当竞争行为

4 尊重合法权益

- 尊重他人合法权益，不得危害他人身心健康，不得侵害他人肖像权、名誉权、荣誉权、隐私权和个人信息权益

5 提高透明度

- 基于服务类型特点，采取有效措施，提升生成式人工智能服务的透明度，提高生成内容的准确性和可靠性

对应《互联网信息服务算法推荐管理规定》

第十三条 算法推荐服务提供者提供互联网新闻信息服务的，应当依法取得互联网新闻信息服务许可，规范开展互联网新闻信息采编发布服务、转载服务和传播平台服务，**不得生成合成虚假新闻信息，不得传播非国家规定范围内的单位发布的新闻信息。**

对应《互联网信息服务深度合成管理规定》

第六条 任何组织和个人不得利用深度合成服务制作、复制、发布、传播法律、行政法规禁止的信息，**不得利用深度合成服务从事危害国家安全和利益、损害国家形象、侵害社会公共利益、扰乱经济和社会秩序、侵犯他人合法权益等法律、行政法规禁止的活动。**

深度合成服务提供者和使用者不得利用深度合成服务制作、复制、发布、传播虚假新闻信息。转载基于深度合成服务制作发布的新闻信息的，应当依法转载互联网新闻信息稿源单位发布的新闻信息。

第五、六条：在技术发展过程中，需要鼓励和推动的方向有哪些？

鼓励创新

创新维度	详情
创新应用	**鼓励生成式人工智能技术在各行业、各领域的创新应用**，生成积极健康、向上向善的优质内容，探索优化应用场景，构建应用生态体系
机构合作	支持行业组织、企业、教育和科研机构、公共文化机构、有关专业机构等在生成式人工智能技术创新、数据资源建设、转化应用、风险防范等方面开展协作
基础创新	**鼓励生成式人工智能算法、框架、芯片及配套软件平台等基础技术的自主创新**，平等互利开展国际交流与合作，参与生成式人工智能相关国际规则制定

推动共享

共享维度	详情
平台建设	推动生成式人工智能基础设施和公共训练**数据资源平台建设**
算力共享	**促进算力资源协同共享**，提升算力资源利用效能
资源开放	**推动公共数据分类分级有序开放**，扩展高质量的公共训练数据资源

安全可信

鼓励采用安全可信的芯片、软件、工具、算力和数据资源

第七条：依法开展预训练、优化训练等训练数据处理活动需满足哪些要求？

生成式人工智能服务提供者（以下称提供者）应当依法开展预训练、优化训练等训练数据处理活动，遵守以下规定：

合法来源

使用具有合法来源的数据和基础模型；

产权保护

涉及知识产权的，不得侵害他人依法享有的知识产权；

信息保护

涉及个人信息的，**应当取得个人同意或者符合法律、行政法规规定的其他情形**；

质量要求

采取有效措施提高训练数据质量，增强训练数据的真实性、准确性、客观性、多样性；

监管要求

《中华人民共和国网络安全法》、《中华人民共和国数据安全法》、《中华人民共和国个人信息保护法》等法律、行政法规的其他有关规定和有关主管部门的相关监管要求。

第八条：技术研发过程中进行数据标注的，需满足哪些具体要求？

炼石 CipherGateway

在生成式人工智能技术研发过程中进行数据标注的，需满足以下要求：

01 规则制定

提供者应当制定符合本办法要求的清晰、具体、可操作的标注规则；

02 质量评估

开展数据标注质量评估，抽样核验标注内容的准确性；

03 人员培训

对标注人员进行必要培训，提升尊法守法意识，监督指导标注人员规范开展标注工作。

对应《互联网信息服务算法推荐管理规定》

第十六条　算法推荐服务提供者应当以显著方式告知用户其提供算法推荐服务的情况，并以适当方式公示算法推荐服务的基本原理、目的意图和主要运行机制等。

对应《互联网信息服务深度合成管理规定》

第十条 深度合成服务提供者应当加强深度合成内容管理，采取技术或者人工方式对深度合成服务使用者的输入数据和合成结果进行审核。

深度合成服务提供者应当建立健全用于识别违法和不良信息的特征库，完善入库标准、规则和程序，记录并留存相关网络日志。

深度合成服务提供者发现违法和不良信息的，应当依法采取处置措施，保存有关记录，及时向网信部门和有关主管部门报告；对相关深度合成服务使用者依法依约采取警示、限制功能、暂停服务、关闭账号等处置措施。

第九条：应当依法承担哪些网络信息内容生产者责任？

提供者

应当依法承担网络信息内容生产者责任，履行网络信息安全义务。

应当与注册其服务的生成式人工智能服务使用者（以下称使用者）签订服务协议，明确双方权利义务

涉及个人信息的

涉及个人信息的，依法承担个人信息处理者责任，履行个人信息保护义务。

第十条：未成年人过分依赖或沉迷相关服务是否需提前防范？

01

适用明确

提供者应当明确并公开其服务的适用人群、场合、用途，指导使用者科学理性认识和依法使用生成式人工智能技术

02

防止沉迷

采取有效措施防范未成年人用户过度依赖或者沉迷生成式人工智能服务

对应《互联网信息服务算法推荐管理规定》

第八条　算法推荐服务提供者应当定期审核、评估、验证算法机制机理、模型、数据和应用结果等，不得设置诱导用户沉迷、过度消费等违反法律法规或者违背伦理道德的算法模型。

第十一条：提供者如何对使用者的输入信息和使用记录承担哪些保护义务？

提供者对使用者的输入信息和使用记录应当依法履行保护义务。

不得收集非必要个人信息

不得非法留存能够识别使用者身份的输入信息和使用记录

不得非法向他人提供使用者的输入信息和使用记录

提供者应当依法及时受理和处理个人关于查阅、复制、更正、补充、删除其个人信息等的请求

对应《互联网信息服务算法推荐管理规定》

第七条　算法推荐服务提供者应当落实算法安全主体责任，建立健全算法机制机理审核、科技伦理审查、用户注册、信息发布审核、数据安全和个人信息保护、反电信网络诈骗、安全评估监测、安全事件应急处置等管理制度和技术措施，制定并公开算法推荐服务相关规则，配备与算法推荐服务规模相适应专业人员和技术支撑。

第二十九条　参与算法推荐服务安全评估和监督检查的相关机构和人员对在履行职责中知悉的个人隐私、个人信息和商业秘密应当依法予以保密，不得泄露或者非法向他人提供。

对应《互联网信息服务深度合成管理规定》

第七条　深度合成服务提供者应当落实信息安全主体责任，建立健全用户注册、算法机制机理审核、科技伦理审查、信息发布审核、数据安全、个人信息保护、反电信网络诈骗、应急处置等管理制度，具有安全可控的技术保障措施。

第十四条　深度合成服务提供者和技术支持者应当加强训练数据管理，采取必要措施保障训练数据安全；训练数据包含个人信息的，应当遵守个人信息保护的有关规定。

第十二条：需要对哪些生成的内容进行标识？

依据《互联网信息服务深度合成管理规定》

提供者

按照

国家互联网信息办公室
中华人民共和国工业和信息化部
中华人民共和国公安部
令
第12号

《互联网信息服务深度合成管理规定》已经2022年11月3日国家互联网信息办公室2022年第21次室务会议审议通过，并经工业和信息化部、公安部同意，现予公布，自2023年1月10日起施行。

国家互联网信息办公室主任 庄荣文
工业和信息化部部长 金壮龙
公安部部长 王小洪
2022年11月25日

第十七条 深度合成服务提供者提供以下深度合成服务，可能导致公众混淆或者误认的，应当在生成或者编辑的信息内容的合理位置、区域进行显著标识，向公众提示深度合成情况：

（一）智能对话、智能写作等模拟自然人进行文本的生成或者编辑服务；

（二）合成人声、仿声等语音生成或者显著改变个人身份特征的编辑服务；

（三）人脸生成、人脸替换、人脸操控、姿态操控等人物图像、视频生成或者显著改变个人身份特征的编辑服务；

（四）沉浸式拟真场景等生成或者编辑服务；

（五）其他具有生成或者显著改变信息内容功能的服务。

深度合成服务提供者提供前款规定之外的深度合成服务的，应当提供显著标识功能，并提示深度合成服务使用者可以进行显著标识。

第十八条 任何组织和个人不得采用技术手段删除、篡改、隐匿本规定第十六条和第十七条规定的深度合成标识。

对图片、视频等生成内容进行标识

*文本内容未要求进行标识

第十三条：提供者如何保障用户正常使用？

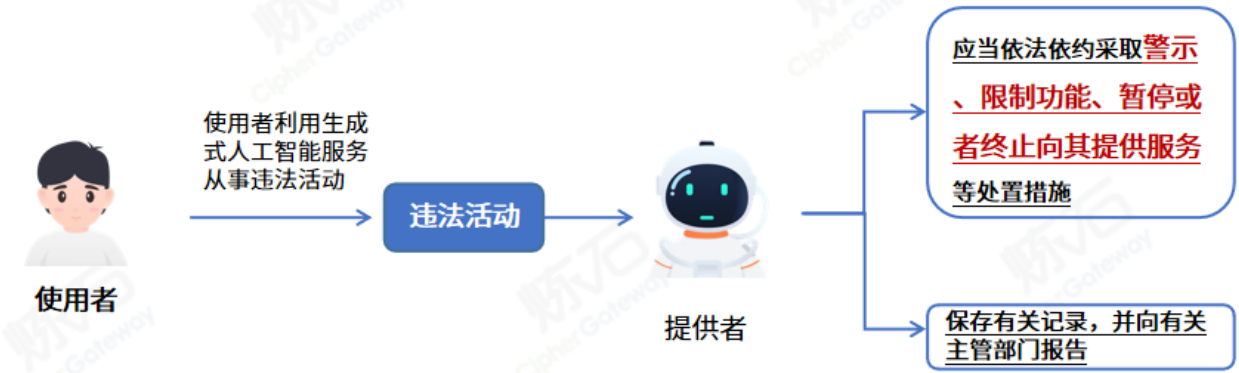
提供者应当在其服务过程中，提供安全、稳定、持续的服务，保障用户正常使用。

第十四条-1：提供者发现违法的内容怎么处理？

提供者发现违法内容的，应当及时采取停止生成、停止传输、消除等处置措施，采取模型优化训练等措施进行整改，并向有关主管部门报告。

第十四、十八条：提供者发现使用者利用相关服务从事违法活动怎么办？

提供者发现使用者利用生成式人工智能服务从事违法活动的，应当依法依约采取警示、限制功能、暂停或者终止向其提供服务等处置措施，保存有关记录，并向有关主管部门报告。



注：使用者发现生成式人工智能服务不符合法律、行政法规和本办法规定的，有权向有关主管部门投诉、举报。

第十五条：提供者如何建立健全用户投诉、举报机制？

炼石 CipherGateway

提供者应当建立健全投诉、举报机制

提供者

建立健全
投诉、举报机制

- 设置便捷的投诉、举报入口
- 公布处理流程和反馈时限，及时受理、处理公众投诉举报并反馈处理结果

对应《互联网信息服务算法推荐管理规定》

第六条　算法推荐服务提供者应当坚持主流价值导向，优化算法推荐服务机制，积极传播正能量，促进算法应用向上向善。

算法推荐服务提供者不得利用算法推荐服务从事危害国家安全和社会公共利益、扰乱经济秩序和社会秩序、侵犯他人合法权益等法律、行政法规禁止的活动，不得利用算法推荐服务传播法律、行政法规禁止的信息，应当采取措施防范和抵制传播不良信息。

对应《互联网信息服务深度合成管理规定》

第六条　任何组织和个人不得利用深度合成服务制作、复制、发布、传播法律、行政法规禁止的信息，不得利用深度合成服务从事危害国家安全和利益、损害国家形象、侵害社会公共利益、扰乱经济和社会秩序、侵犯他人合法权益等法律、行政法规禁止的活动。

深度合成服务提供者和使用者不得利用深度合成服务制作、复制、发布、传播虚假新闻信息。转载基于深度合成服务制作发布的新闻信息的，应当依法转载互联网新闻信息稿源单位发布的新闻信息。

第十二条　深度合成服务提供者应当设置便捷的用户申诉和公众投诉、举报入口，公布处理流程和反馈时限，及时受理、处理和反馈处理结果。

第十六条：各部门职责有哪些?

炼石 CipherGateway

职责主体	具体职责
网信、发展改革、教育、科技、工业和信息化、公安、广播电视、新闻出版等部门	依据各自职责**依法加强**对生成式人工智能服务的管理
国家**有关主管**部门	针对生成式人工智能技术特点及其在有关行业和领域的服务应用，完善与创新发展相适应的**科学监管方式**，制定相应的**分类分级监管规则或者指引**

对应《互联网信息服务算法推荐管理规定》

第三条 国家网信部门负责统筹协调全国算法推荐服务治理和相关监督管理工作。国务院电信、公安、市场监管等有关部门依据各自职责负责算法推荐服务监督管理工作。

第二十三条 **网信部门会同电信、公安、市场监管等有关部门建立算法分级分类安全管理制度**，根据算法推荐服务的舆论属性或者社会动员能力、内容类别、用户规模、算法推荐技术处理的数据重要程度、对用户行为的干预程度等对算法推荐服务提供者**实施分级分类管理**。

对应《互联网信息服务深度合成管理规定》

第三条 国家网信部门负责统筹协调全国深度合成服务的治理和相关监督管理工作。国务院电信主管部门、公安部门依据各自职责负责深度合成服务的监督管理工作。

第十七条-1：需要安全评估的具体情形和评估事项？

炼石 CipherGateway

提供**具有舆论属性或者社会动员能力的生成式人工智能服务**的，应当**按照国家有关规定开展安全评估**

安全评估：依据国家网信办和公安部2018年11月联合发布的**《具有舆论属性或社会动员能力的互联网信息服务安全评估规定》**，向国家网信部门**申报安全评估**

需要安全评估的情形	安全评估事项
（一）具有舆论属性或社会动员能力的信息服务上线，或者信息服务增设相关功能的； （二）使用新技术新应用，使信息服务的功能属性、技术实现方式、基础资源配置等发生重大变更，导致舆论属性或者社会动员能力发生重大变化的； （三）用户规模显著增加，导致信息服务的舆论属性或者社会动员能力发生重大变化的； （四）发生违法有害信息传播扩散，表明已有安全措施难以有效防控网络安全风险的； （五）地市级以上网信部门或者公安机关书面通知需要进行安全评估的其他情形。	（一）确定与所提供服务相适应的安全管理负责人、信息审核人员或者建立安全管理机构的情况； （二）用户真实身份核验以及注册信息留存措施； （三）对用户的账号、操作时间、操作类型、网络源地址和目标地址、网络源端口、客户端硬件特征等日志信息，以及用户发布信息记录的留存措施； （四）对用户账号和通讯群组名称、昵称、简介、备注、标识，信息发布、转发、评论和通讯群组等服务功能中违法有害信息的防范处置和有关记录保存措施； （五）个人信息保护以及防范违法有害信息传播扩散、社会动员功能失控风险的技术措施； （六）建立投诉、举报制度，公布投诉、举报方式等信息，及时受理并处理有关投诉和举报的情况； （七）建立为网信部门依法履行互联网信息服务监督管理职责提供技术、数据支持和协助的工作机制的情况； （八）建立为公安机关、国家安全机关依法维护国家安全和查处违法犯罪提供技术、数据支持和协助的工作机制的情况。

第十七条-2：算法备案如何执行落地？

炼石 CipherGateway

按照《互联网信息服务算法推荐管理规定》履行算法备案和变更、注销备案手续

注：自《互联网信息服务算法推荐管理规定》实施以来，国家网信办于2022年8月、10月及2023年1月先后三次公布了《境内互联网信息服务算法备案清单》，发布了200余项算法备案，其中包括了多家头部互联网公司的算法。

算法备案主体	制度要求	算法备案要求	处罚
1.为公众舆论提供表达渠道的服务提供者； 2.具有发动社会公众从事特定活动的互联网信息服务商。	1.算法备案主体进行算法备案填报时应当对落实算法安全主体责任的情况进行说明； 2.在进行算法备案前，企业应设置算法安全机构，建立完善的公司内部规章制度。	具有舆论属性或者社会动员能力的算法推荐服务提供者应当在提供服务之日起十个工作日内通过互联网信息服务算法备案系统填报服务提供者的名称、服务形式、应用领域、算法类型、算法自评估报告、拟公示内容等信息，履行备案手续。 算法推荐服务提供者的备案信息发生变更的，应当在变更之日起十个工作日内办理变更手续。 算法推荐服务提供者终止服务的，应当在终止服务之日起二十个工作日内办理注销备案手续，并作出妥善安排。	具有舆论属性或者社会动员能力的算法推荐服务提供者通过隐瞒有关情况、提供虚假材料等不正当手段取得备案的，由国家和省、自治区、直辖市网信部门予以撤销备案，给予警告、通报批评；情节严重的，责令暂停信息更新，并处一万元以上十万元以下罚款。 具有舆论属性或者社会动员能力的算法推荐服务提供者终止服务未按照本规定第二十四条第三款要求办理注销备案手续，或者发生严重违法情形受到责令关闭网站、吊销相关业务许可证或者吊销营业执照等行政处罚的，由国家和省、自治区、直辖市网信部门予以注销备案。

注：部分内容来源于CCIA数据安全工作委员会公众号

第十九条：有关主管部门依据职责对相关服务开展监督检查的要求有哪些？

炼石 CipherGateway

《生成式人工智能服务管理暂行办法》

监督检查

有关主管部门依据职责对生成式人工智能服务开展监督检查，提供者应当依法予以配合，按要求对训练数据来源、规模、类型、标注规则、算法机制机理等予以说明，并提供必要的技术、数据等支持和协助。

依法保密

参与生成式人工智能服务安全评估和监督检查的相关机构和人员对在履行职责中知悉的国家秘密、商业秘密、个人隐私和个人信息应当依法予以保密，不得泄露或者非法向他人提供。

《互联网信息服务深度合成管理规定》

监督检查

第二十一条 网信部门和电信主管部门、公安部门依据职责对深度合成服务开展监督检查。深度合成服务提供者和技术支持者应当依法予以配合，并提供必要的技术、数据等支持和协助。

《互联网信息服务算法推荐管理规定》

依法保密

第二十九条 参与算法推荐服务安全评估和监督检查的相关机构和人员对在履行职责中知悉的个人隐私、个人信息和商业秘密应当依法予以保密，不得泄露或者非法向他人提供。

第二十、二十一条：违反相关规定的服务提供者将面临哪些处罚？

情况说明	主管部门	违规说明	处罚详情
境外向境内提供服务	国家网信部门	对来源于中华人民共和国境外向境内提供生成式人工智能服务不符合法律、行政法规和本办法规定的	应当**通知有关机构采取技术措施和其他必要措施予以处置**
境内服务提供者	有关主管部门	提供者违反本办法规定的	**依照《中华人民共和国网络安全法》、《中华人民共和国数据安全法》、《中华人民共和国个人信息保护法》、《中华人民共和国科学技术进步法》等法律、行政法规的规定予以处罚**
		法律、行政法规没有规定的	**依据职责予以警告、通报批评，责令限期改正；拒不改正或者情节严重的，责令暂停提供相关服务**
		构成违反治安管理行为的	依法给予治安管理处罚
		构成犯罪的	依法追究刑事责任

第二十二、二十三、二十四条：明确了哪几种具体的用语定义？

类别	用语/许可情况	详情
用语定义	生成式人工智能**技术**	是指具有文本、图片、音频、视频等内容生成能力的模型及相关技术
	生成式人工智能**服务提供者**	是指利用生成式人工智能技术提供生成式人工智能服务（包括通过提供可编程接口等方式提供生成式人工智能服务）的组织、个人
	生成式人工智能**服务使用者**	是指使用生成式人工智能服务生成内容的组织、个人
行政许可	法律、行政法规规定提供生成式人工智能服务**应当取得相关行政许可的**	提供者**应当依法取得许可**
	外商投资生成式人工智能服务	**应当符合外商投资相关法律、行政法规的规定**
施行日期	本办法自**2023年8月15日**起施行	

行业数据安全建设实践

安全需求：“风险与合规”共同驱动数据安全建设

炼石 CipherGateway

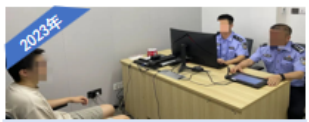
内在风险：数据风险如影随形伴生数据处理

数据安全产业发展阶段性（先发展再治理，迎来拐点）、行业特殊性（经济学外部效应带来密集法律法规），塑造实战防护与合规需求

江苏某地方农商银行因违反网络安全与数据安全管理规定等被警告，并被罚款114.5万元。

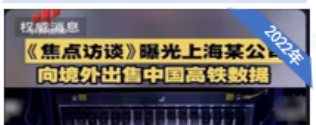
某国有股份制商业银行因数据安全管理等问题被金监局罚款160万元

上海市网信办会同市市场监管局、市国资委依法联合约谈**某大型地铁集团**及三家运营企业



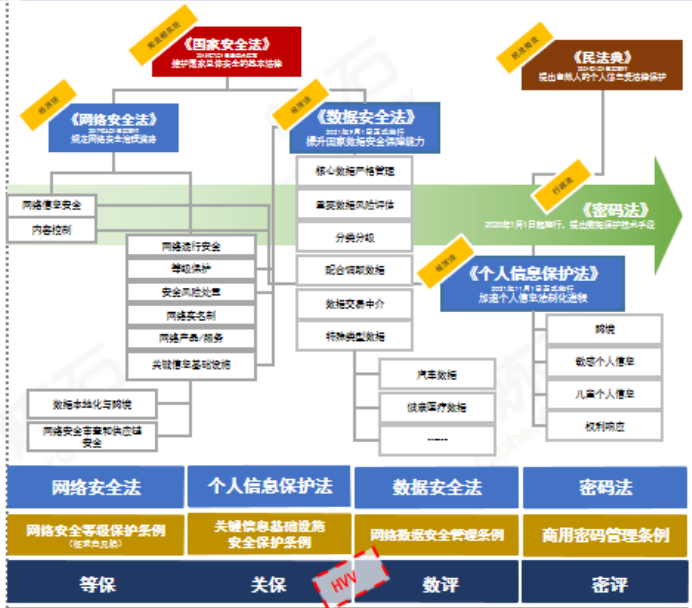
浙江某科技公司为浙江县级市政府开发运维信息管理系统，私自迁移数据，造成严重数据泄露，被罚款100万元。

南昌某高校3万余条师生个人信息数据在境外互联网上被公开售卖，被南昌公安网安部门罚款85万元。

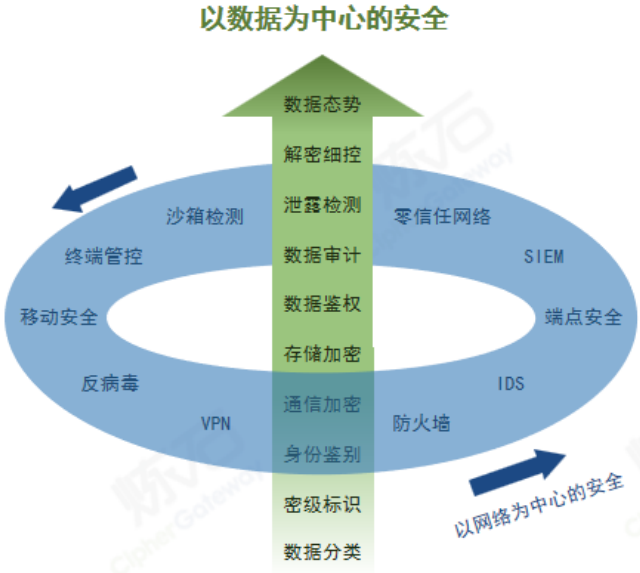


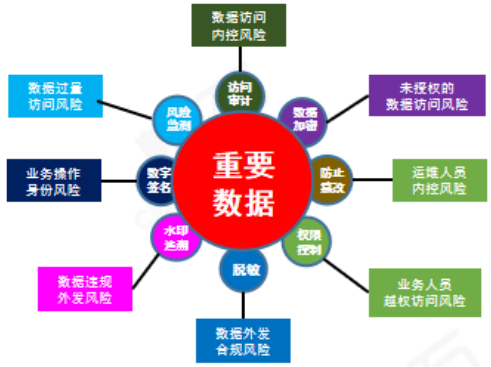
国家安全机关破获我国**首例**涉及高铁运行安全的**危害国家安全类**案件

外部合规：“五法一典”过程与结果双合规



落地痛点：安全建设从网络到数据，安全和业务纽结缠绕难以改造

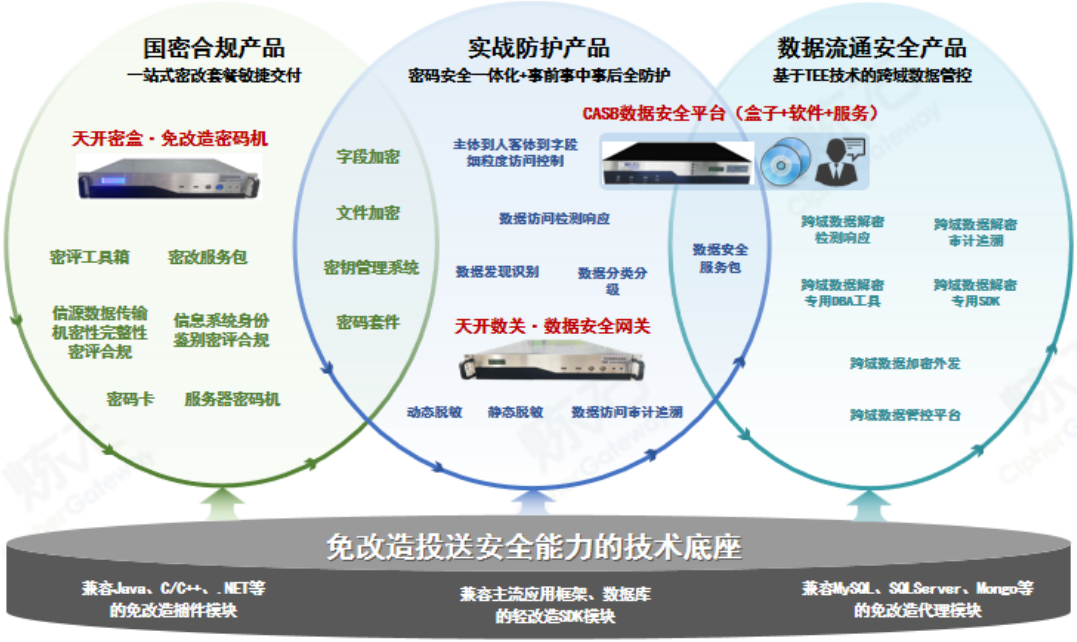




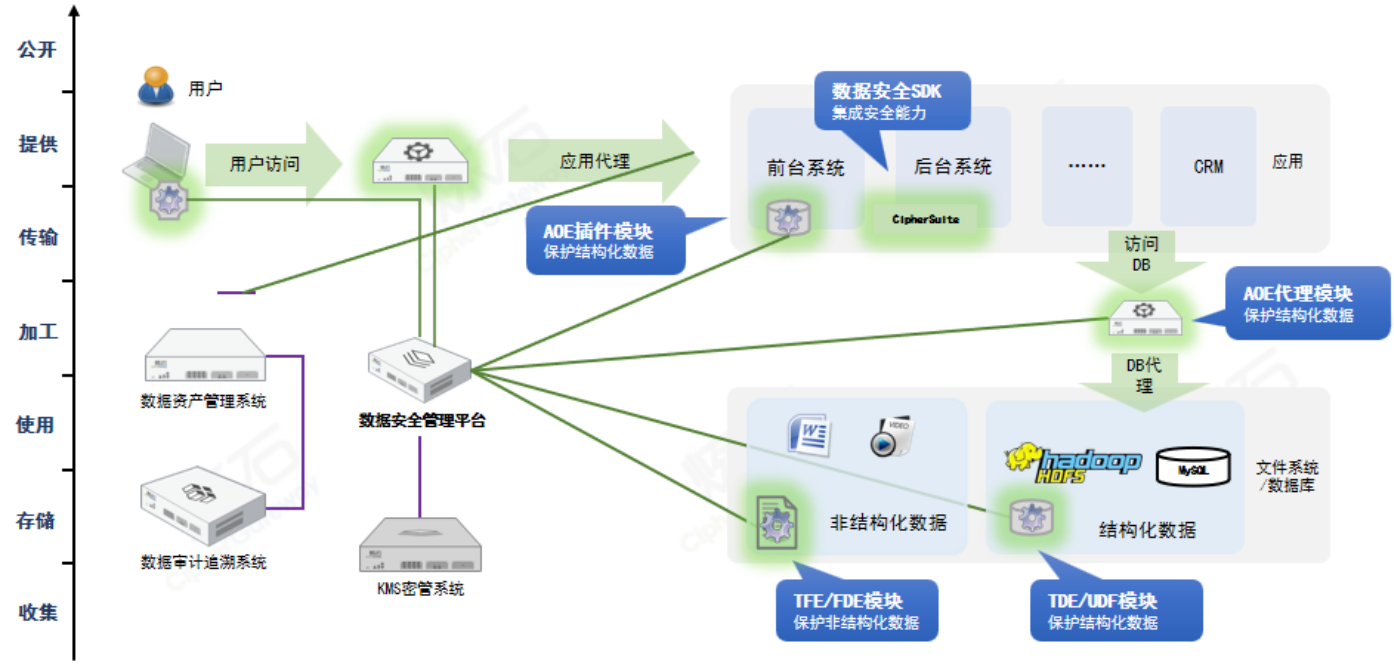
数据安全建设面临“两难”：

- 数据保护体现为**面向应用的功能型安全需求**，要在应用中融合实现，**但改造存量应用成本高、风险大、周期长，新应用的业务开发与安全排期也不匹配，安全机制总是滞后或缺失**
- 而数据保护不落实，则合规风险越来越大、因泄露导致的业务风险会快速累积

产品体系：免改造平台灵活交付多重安全能力，易部署易扩展



敏捷部署：旁路平台结合主路控制面，免改造应用落地保护



场景之一：存储加密、使用解密/脱敏

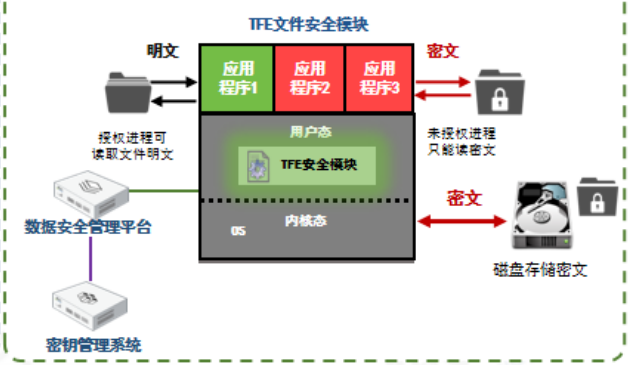
炼石 CipherGateway

存储 使用 加工 传输 提供

结构化数据

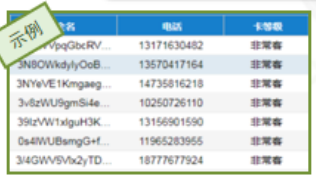
支持ABAC的访问控制策略，实现用户与字段文档级防护

非结构化数据

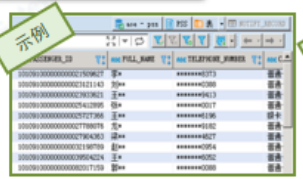


不影响业务人员使用、不影响DBA运维

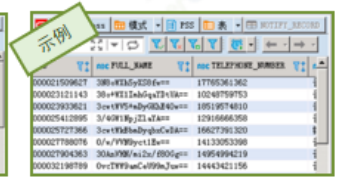
应用查看脱敏后数据



应用看到密文



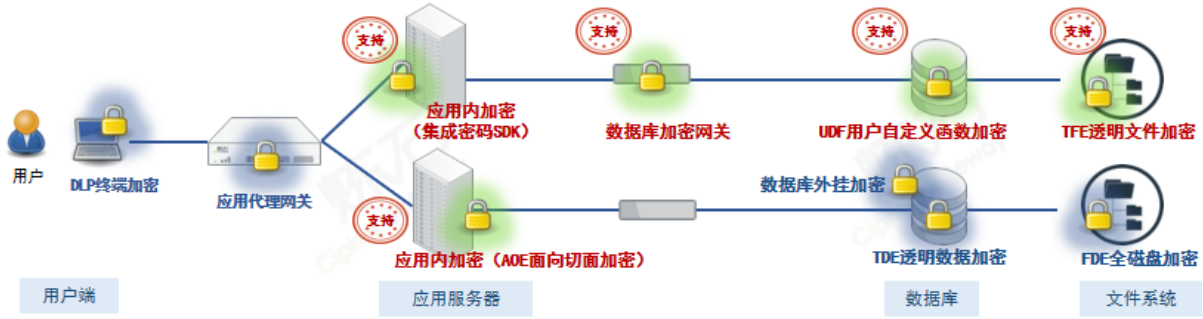
DBA工具查看脱敏数据



DBA工具查看密文

一站式支持多技术路线，结合业务场景交付最优组合

炼石 CipherGateway



加密技术	部署位置	加密粒度	性能	防DBA	DB存储过程密文计算	实施成本	可靠性
DLP终端加密	终端	文件	中	/	/	中	中
应用代理网关	终端-应用服务器之间	字段/文件	中	支持	影响	高	低
应用内加密（集成密码SDK）	应用服务器	字段/文件	高	支持	影响	高	高
应用内加密（AOE面向切面加密）	应用服务器	字段	高	支持	影响	低	高
数据库加密网关	应用服务器-数据库之间	字段	中	支持	影响	低	中
数据库外挂加密	数据库	字段	低	支持	影响	中	中
TDE透明数据加密	数据库	字段/表空间	高	不支持	不影响	低	中
UDF用户自定义函数加密	数据库	字段	高	不支持	不影响	高	高
TFE透明文件加密	文件系统	文件	中	不支持	不影响	低	高
FDE全磁盘加密	文件系统	磁盘/卷	高	不支持	不影响	低	高

产品价值：将安全能力投送到真实的数据处理流程中

免改造技术对流动数据的高覆盖率识别与控制，实现安全与业务有机融合

国密合规场景

一站式密改套餐敏捷交付

密评要点解析

炼石 CipherGateway

- 最贴合业务
- 业务风险影响范围广，改造难度大
- 涉及核心的高风险项，且分值最高

- 属于密码应用层，含高风险项最多
- 密改落地由用户执行，无法绕过
- 所需密码产品会随着信息系统新增或变化，而扩容或调整

- 属于密码支撑层
- 密改落地不一定由用户执行，有可能属于IDC托管方、云服务商等
- 整体被评，可复用到被承载的每个信息系统

密码测评规则

依据	GB/T 39786-2021 《信息安全技术 信息系统密码应用基本要求》
通过要求	60分以上，无高风险
测评规则	打分制，满分100，共41项检测项，包含8个方面
典型问题	无合规的密码设备，无管理制度
评估周期	重要网络与信息系统（关基、等保三级、政务等）每年至少一次

高风险项，一票否决

- 实体间未通过密码技术进行身份鉴别（网络和通信：无缓解措施）
- 未采用密码技术保证通信过程中重要数据的机密性
- 未采用密码技术建立安全的信息传输通道，如SSLVPN
- 未采用密码技术保证重要数据在存储过程中的机密性（无缓解措施）
- 未采用密码技术保证重要数据在存储过程中的完整性
- 未采用密码技术保证重要数据在传输过程中的机密性
- *未采用密码技术提供数据原发证据和数据接收证据（法律责任场景）*
- *未建立密码相关管理制度*
- *未制定密码应用解决方案、实施方案和应急处置方案，未通过评审*
- *未采用密码技术对从外部连接到内部的设备进行认证（第四级）*

密码合规要点（以密码应用三级系统为例

物理和环境	身份鉴别	门禁记录	监控记录	密码服务资质	密码产品认证			10分
网络和通信	身份鉴别	数据的完整性和机密性	访问控制信息完整性	设备认证	密码服务资质	密码产品认证		20分
设备和计算	身份鉴别	加密通道	访问控制和安全标记的完整性	日志和程序完整性	真实性	密码服务资质	密码产品认证	10分
应用和数据	身份鉴别	传输和存储的机密性	完整性	不可否认性	密码服务资质	密码产品认证		30分
管理制度	密码人员管理	密钥管理	建设运行	应急处置	密码软硬件	介质管理等制度		
人员管理	角色	职责	权限	审计机制	培训与考核	保密协议		
建设运行	密码应用方案	实施方案						30分
应急处置	应急处置方案	及时上报						

70分密改套餐

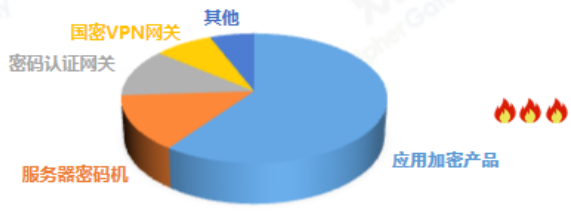
测评层面	测评单元	备注	满分	得分	密码产品
物理和环境安全（10分）	身份鉴别	高风险	4.17	0.00	高风险可缓解
	电子门禁记录数据存储完整性		2.92	0.00	
	视频监控记录数据存储完整性		2.92	0.00	
网络和通信安全（20分）	身份鉴别	高风险（无缓解）	5.88	5.88	IPSec VPN安全网关 SSL VPN安全网关
	通信数据完整性		4.12	4.12	
	通信过程中重要数据的机密性	高风险	5.88	5.88	
	网络边界访问控制信息的完整性		4.12	4.12	
设备和计算安全（10分）	身份鉴别	高风险	2.08	2.08	SSL VPN安全网关
	远程管理通道安全	高风险	2.08	2.08	
	系统资源访问控制信息完整性		1.46	1.46	
	重要信息资源安全标记完整性		1.46	1.46	密钥管理与数据加密及认证平台
	日志记录完整性		1.46	1.46	
	重要可执行程序完整性、重要可执行程序来源真实性		1.46	0.00	
应用和数据安全（30分）	身份鉴别	高风险	4.05	4.05	密钥管理与数据加密及认证平台 服务器密码机 软件密码模块 密码卡
	访问控制信息完整性		2.84	2.84	
	重要信息资源安全标记完整性		2.84	2.84	
	重要数据传输机密性	高风险	4.05	4.05	
	重要数据存储机密性	高风险（无缓解）	4.05	4.05	
	重要数据传输完整性		4.05	4.05	
	重要数据存储完整性	高风险	4.05	4.05	
	不可否认性	高风险	4.05	0.00	仅限法律责任场景
安全管理（30分）	按得分80%计算		30	24	专业管理制度模板
合计			100	78.49	

* 参考密评联委会《商用密码应用安全性评估量化评估规则（2023版）》

密改四件套，轻松过密评

国密VPN网关产品	密码认证网关产品	应用加密网关产品
解决网络和通信20分、缓解高风险	解决应用和数据5分、缓解高风险	应用和数据21分、高风险，设备和计算4分

服务器密码机
提供基础加解密算力与密管能力



高性能：PCT专利保护的高速国密实现，不影响业务效率

炼石 CipherGateway

基于PKI的数字信封加密 | 协同签名身份鉴别

密钥派生接口 | EVP国密接口 | FPE格式保留加密

Java | C# | python | Ruby

Programming | C++ | GO

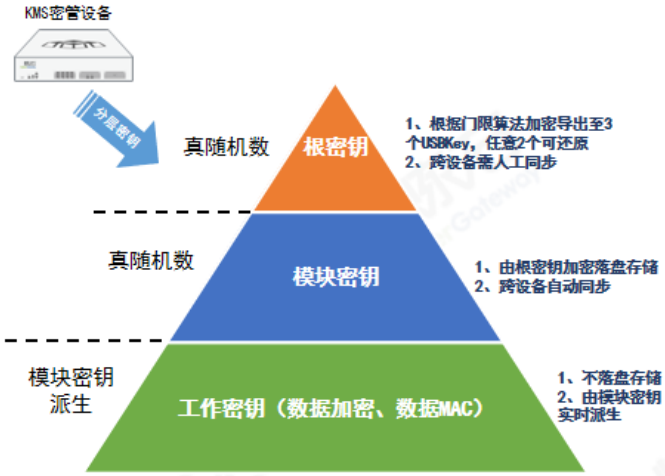
SM2 | SM3 | SM4 | SM9 | ZUC

Windows 10 | Linux | Mac OS | android | iOS

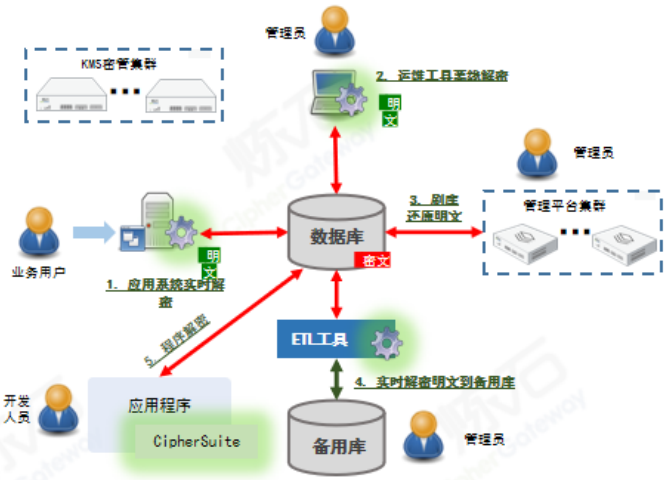
基于SM4的保留格式加密（FPE）

加密10亿条手机号，仅耗时20秒

高可靠：专业密钥管理、实时离线解密，保障高可用

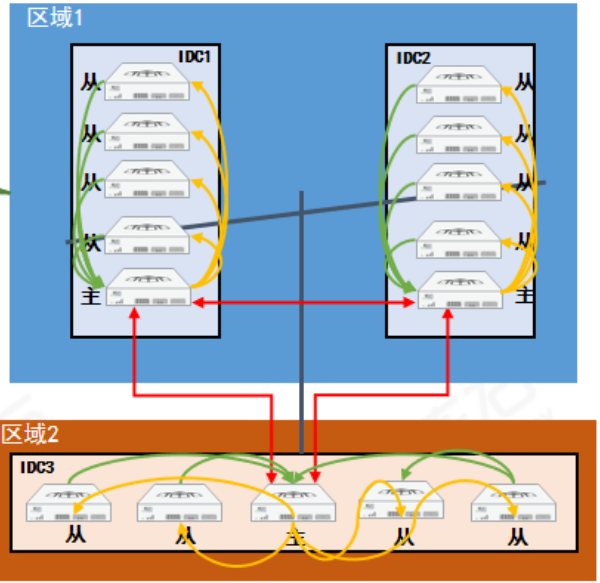


- 密管设备**支持双机热备、两地三中心等高可用模式**，根密钥明文不出密码卡，模块密钥明文不出密管设备
- **工作密钥安全从平台分发到数据控制面模块，模块可缓存工作密钥后独立运行**，平台不需要实时在线
- 保守策略下，当重启应用服务后，模块才需从平台更新密钥



- 密码算法是标准，密钥管理机制可靠，**只要密文数据还在，一定可解密**
- 常态状态下，支持应用系统解密、支持DBA运维工具直接解密
- 应急情况下，支持刷库还原明文、支持实时解密明文到备用库
- 极端情况下，用运维工具或程序可以兜底解密

高可用：模块健壮可诊断、平台“两地三中心”



【模块可离线运行】

- 模块从平台获取规则后可缓存在内存中，当平台不可用也可正常运行
- 仅当模块所在服务重启时才需要从平台集群重取规则

【模块强化健壮性】

- 以AOE插件为例，本身作为应用的一部分，无单独进程，不会死进程
- 在没有匹配规则状态下，模块默认采用透传模式，最大限度优先业务
- 模块经过多年研发打磨、运行实践，兼容广泛苛刻的企业级应用场景

【模块可诊断】

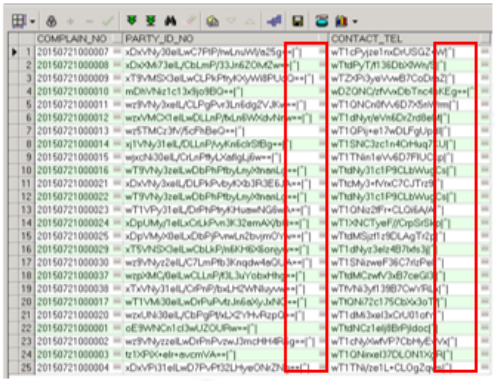
- 模块自身可记录环境、启动流程、异常事件等日志，以及告警功能
- 针对Java等应用环境支持不停机实时诊断机制，排查潜在故障成因

- 两地三中心各5套安全管理平台（以及KMS）组成集群，三中心的主之间实时双向同步数据，在同一中心内主从间同步数据

工业级：加密实现机制兼容复杂场景

炼石 CipherGateway

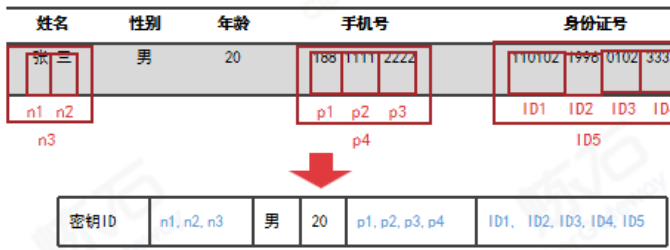
明密文标识确保加解密可靠



- 炼石产品加密后，会在密文中添加密文标识，可保证明文不会被二次解密，密文不会被二次加密
- 在加密功能上线过程中，不可避免会出现明密文混合的情况，可通过明密文标识将明文和密文区分开

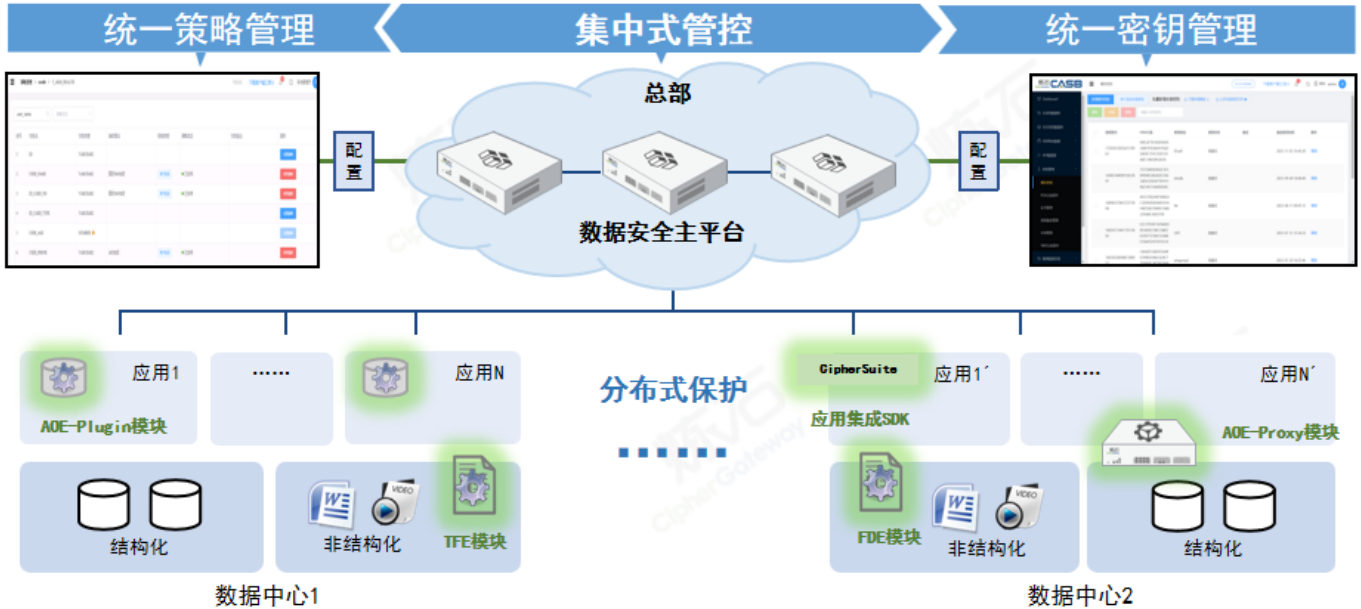
支持数据加密后的密文态查询

类型	规则	示例
a	手机号型	按照3、4、4拆分加密
b	身份证号型	按照6、4、4、4拆分加密
c	住址型	按照3、3、其余拆分加密
e	其他	------



- 支持密态数据库的模糊查询操作。对加密字段做特定处理，先将加密字段按照规则拆解，然后对拆解后部分分别加密处理，之后将各部分密文合成一个字段（拆解的规则可由应用定制化，通过策略管理来进行配置）
- 限制条件
  - 支持应用对加密策略进行定制，但加密策略确定后原则上不建议更改
  - 为支持模糊查询会对加密字段的长度进行扩张

好用好管：集中式管控，分布式保护，分阶段上线



数据安全实战防护场景

事前事中事后全防护

传统密服平台面临挑战

传统密码服务平台（CSP）

- 【易用】从基础密码裸算法能力升级到服务化密码能力，易于集成
- 【提效】实现从密码烟囱式硬件到密码资源池化，适应了云化趋势
- 【好管】从功能上支持分层次、多租户等管理模式，方便集中部署

数据实战化防护挑战

传统CSP仅侧重于密码技术本身应用，而加密仅把明文安全转移为密钥安全，如果缺乏事前、事中、事后等全方位数据保护手段，在真实业务场景中难以输出有效安全防护能力

密改的应用改造难题

传统CSP以供应加解密算力为主，在集成到现有信息系统时需要进行大量二次开发工作，应用开发改造风险大、成本高、周期长，在密评推广普及后日益突出上升为主要矛盾

安全新合规支撑不足

《数据安全法》《个人信息保护法》《网络数据安全管理条例》《关基条例》等对数据安全的合规要求日趋严格，相应的关保、各类数据安全合规等新需求对密码提出明确要求，而传统CSP在满足新合规方面缺乏灵活性，难以支撑落地

下一代密服平台打造关键能力

NGCSP, Next Generation Cryptography Service Platform

下一代密码服务平台基于免改造创新技术，实现密码能力从支撑层到真正应用层的跨越，有效解决“密评密改”推广普及后的落地难题，是顺应数字化发展趋势，基于创新切面安全技术构建的综合服务平台。

构建实战能力

通过免改造的整合优势，将密码能力与数据安全能力紧密结合，形成了事前脱敏/加密/访问控制、事中检测响应、事后审计追溯的全方位数据安全实战防护体系

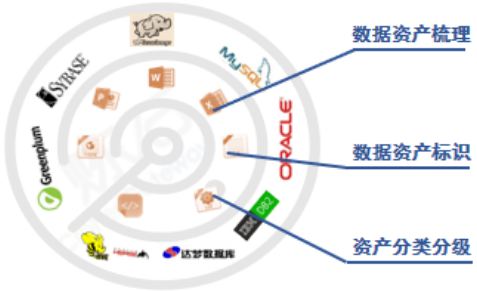
贴合密改重心

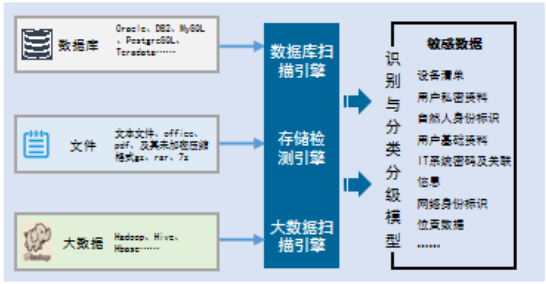
通过免改造技术优势，轻松接入现有信息系统，提供贴近业务场景的密码解决方案，实现从密码支撑层向真正密码应用层的跨越

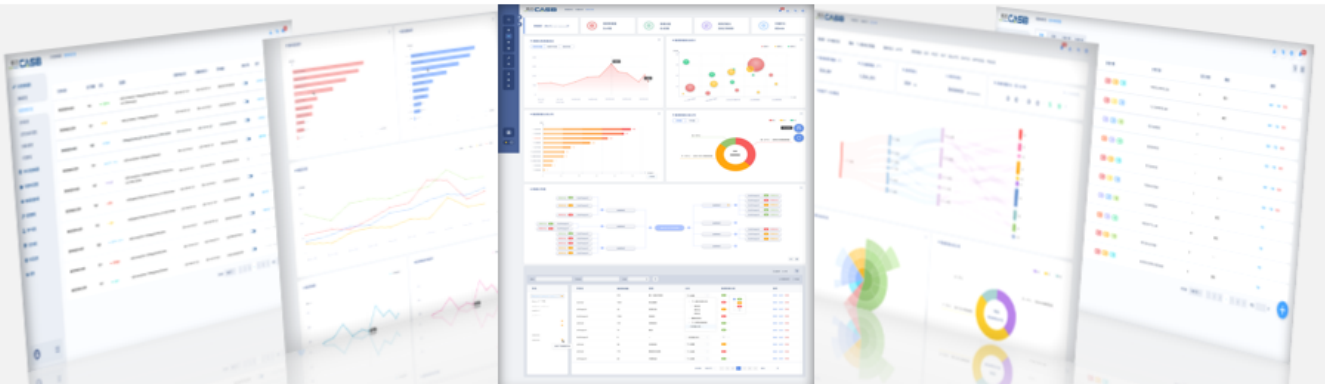
统筹数安合规

满足密评、关保、等保、数评等合规要求、数据安全两法一条例推动的数据加密合规要求、工作秘密防护等数据安全新合规融合

数据可见：结合法规要求，落地数据分类分级



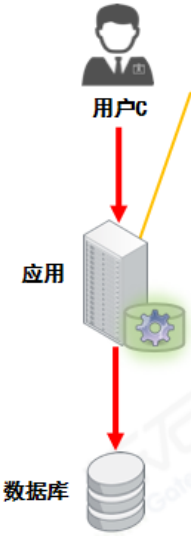




事前防护：最小化敏感数据访问范围

- 可适配应用系统自带身份管理或统一IAM系统，实现结合用户身份的动态脱敏
- 可支持ABAC基于属性的访问控制等细粒度授权

事中检测响应：业务人员风险操作的识别处置



示例

用户C 20161221 23:43:22 下载文件 产品库/详细设计/ 作动器总体设计图（最终）

行为不符合日常工作时间　行为异常　执行阻断

场景举例

某员工用户C有权使用单位的PLM系统，并能访问到某Z型产品中的设计数据。偶然一次，此人在正常工作时间内，批量下载全套的产品数据，此时数据控制面的实时监控功能捕捉到此异常行为，会立即执行阻断，并将该事件进行记录和报警。

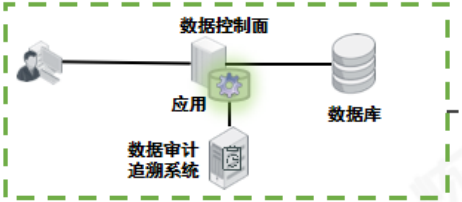
- 免改造控制面对所有流经的业务操作进行监测，结合当前的业务上下文，判断行为是否符合正常的行为模式
- 如果发现是异常操作和不合规操作，则进行告警甚至对操作阻断，以防止恶意操作带来的数据泄露和系统破坏

示例

客户信息安全内控增强

- 客服人员每天需要查询高价值的大客户信息，以便提供服务，但存在过量访问的内控风险，而业务应用往往缺失细粒度安全机制
- 通过免改造控制面的安全内控增强，可实现客服人员每天访问大客户信息的阈值，比如100条以内可正常访问，超过100条系统进行告警，超过200条会触发阻断（通过流程审批后才能访问）

事后审计：可定责到人的访问审计追溯

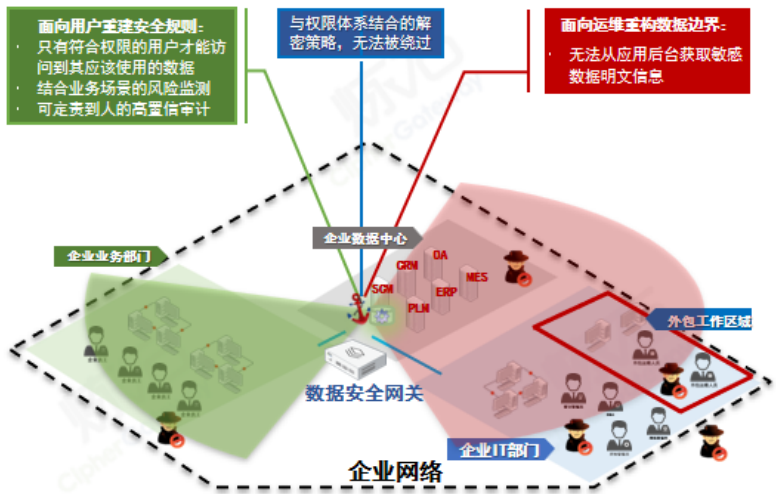


记录所有访问敏感数据的行为

包括账号、时间、IP、会话、操作、对象、耗时、结果等等内容，形成包含终端用户身份的SQL日志。用户可通过日志查询数据的访问情况，在事故发生后精确溯源

操作时间	操作者	行为	操作对象	操作对象ID	操作对象名称	操作对象所有者	状态	是否告警	告警类型	附件属性	客户端IP	日志签名
2018-02-01 16:13:25	kunu	下载	附件	09_9	09_9	kunu	成功	否	否	279568	172.16.10.98	MEUCIQC7L...
2018-02-01 16:13:25	kunu	下载	附件	01_1	01_1	kunu	成功	否	否	213008	172.16.10.98	MEUCIApAW...
2018-02-01 16:13:25	kunu	下载	附件	03_3	03_3	kunu	成功	否	否	95760	172.16.10.98	MEQCIOYgO...
2018-02-01 16:13:24	kunu	下载	附件	01_13	01_13	kunu	成功	否	否	96784	172.16.10.98	MEUCIQDYwetNGDYAdwKlp64s89bMw9OhaWUvWMT
2018-02-01 16:13:24	kunu	下载	附件	02_2	02_2	kunu	成功	否	否	376336	172.16.10.98	MEUCIQDYw...
2018-02-01 16:13:24	kunu	下载	附件	04_4	04_4	kunu	成功	否	否	163344	172.16.10.98	MEUCI89FG...

实战防护：密码安全一体化，免改造打造数据防绕过保护



控制面投送多样安全能力，**优先落地数据加密脱敏**

技术演进：下一代密码服务平台是叠加演进，而非替代演进

炼石 CipherGateway

合规遵循

四法：网络安全法 | 密码法 | 数据安全法 | 个人信息保护法

四例：网络安全等级保护条例(征) | 关键信息基础设施安全保护条例 | 商用密码管理条例 | 网络数据安全管理条例

四规：等保 | 关保 | 密评 | 数评

业务应用

业务应用系统

数据收集	数据预处理	数据传输	数据存储	数据使用	数据加工	数据提供公开	其他基础技术
关系型数据库离线/实时收集	数据清洗	数据库对库同步	单机存储	实时计算	ROLAP / HOLAP	可视化库	资源管理
日志数据离线/实时收集	数据集成	中间件传输	分布式存储	离线计算	MOLAP / OLTP	可视化平台	任务调度系统
网页数据实时收集	数据转换	接口传输/文件传输				可视化工具	全文检索引擎

数据安全能力

事前		事中		事后
数据资产目录	字段级加密	业务威胁检测	威胁响应接口	数据访问留存
数据识别管理	文件级加密	API数据风险	内容/接入阻断	数据库/文件访问审计
自动化分类分级	数据脱敏	数据访问检测	攻击定位	数据访问溯源
数据标注	TDF可信数据格式	数据流转预警	数据威胁情报	数据水印

基础密码能力

通用密码服务	支撑密码服务	密钥管理
加解密服务	数字证书服务	对称密钥管理
随机数服务	电子签章服务	非对称密钥管理
数字信封服务	身份认证服务	标识密钥管理
签名验签服务	安全通信服务	

密码资源池化

密码资源池化：设备管理 | 资源编排 | 资源调度 | 弹性扩容 | 灾备恢复 | 资源监控

密码卡 | 密码机 | 云密码机 | 密钥管理服务器 | 协同签名服务器 | 网络 | 服务器 | 数据库

无缝升级：支持纳管已建密码设备，敏捷落地免改造数据安全

炼石 CipherGateway

下一代密码服务平台（NGCSP）通过统一纳管密码资源池、构建实战防护体系、深度融合合规要求、云原生架构与智能化服务，实现了密码能力从支撑层到应用层的跨越。它不仅提升了信息系统的整体安全水平，还为政企用户提供了高效、安全、合规的密码服务，助力数字化时代的全面安全防护。

统一纳管密码资源池 无缝落地免改造加密

- **统一纳管与调度**：支持对用户已有的传统密码设备和系统进行统一纳管、调度和监控，兼容多厂商的密码设备和系统，消除“密码烟囱”问题，实现密码资源的集中化管理。
- **无缝升级与免改造**：通过免改造技术，无缝对接现有信息系统，无需对业务系统进行大规模改造，即可实现密码能力的升级和落地，降低实施成本和时间。
- **弹性扩展与资源池化**：将密码硬件设备资源池化，支持弹性扩展，满足云计算和大数据环境下的动态需求，提升密码服务的灵活性和可扩展性。

构建全方位数据安全 实战防护体系

- **事前防护**：通过数据脱敏、访问控制等技术，确保数据在存储和传输过程中的机密性和完整性。
- **事中检测与响应**：实时监控数据访问行为，快速检测异常并响应，防止数据泄露或篡改。
- **事后审计与追溯**：提供全面的审计日志和追溯功能，确保数据操作透明可查，满足合规要求。

深度融合数据安全 新合规要求

- **密评与等保合规**：支持密码应用安全性评估（密评）、关键信息基础设施保护（关基）和等保合规要求。
- **数据安全两法一条例**：满足《网络安全法》、《数据安全法》和《个人信息保护法》的合规要求，确保数据加密和隐私保护符合法规。
- **工作秘密防护**：提供针对工作秘密的专项防护措施，确保敏感信息安全。

云原生架构与智能化 密码服务

- **云原生支持**：采用容器化和微服务架构，适应云计算环境，提升系统的弹性和可扩展性。
- **智能化管理**：通过自动化工具实现密钥的生成、分发和更新，减少人工干预，提升密码管理的效率和安全性。
- **自适应安全策略**：根据实时威胁情报动态调整安全策略，确保系统能够应对不断变化的安全威胁。

数据流通场景

基于TEE技术的跨域数据管控

外发痛点：数据“可见即可转”，共享失控

数据已成为企业间可以交易的商品，但在共享后存在“数据失控”问题困扰着数据提供方：

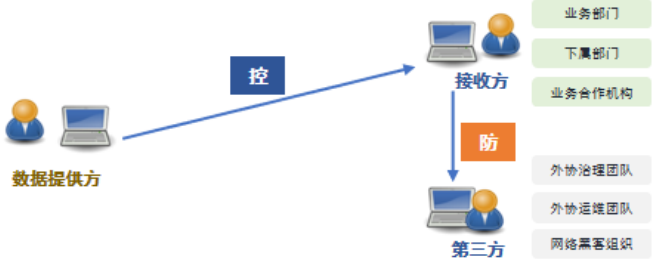
- 数据的接收方有没有对数据进行滥用？
- 对方有没有再次将数据外发给第三方？
- 这些数据在对方的有效期如何控制？

问题实质是，**数据提供方仅赋予接收方数据使用权而非所有权，并且要按照约定来限制对方的使用范围。**

现行做法是双方签订协约实现约束，但**数据的所有权从技术上是失控的**。数据不得不共享的前提下，**必须从技术上实现对接收方使用数据的行为进行控制**，防止数据滥用、防止数据泄露给第三方，维护数据提供方的利益。

跨域数据：基于TEE的TDF技术，轻改造落地数据“可见不可转”

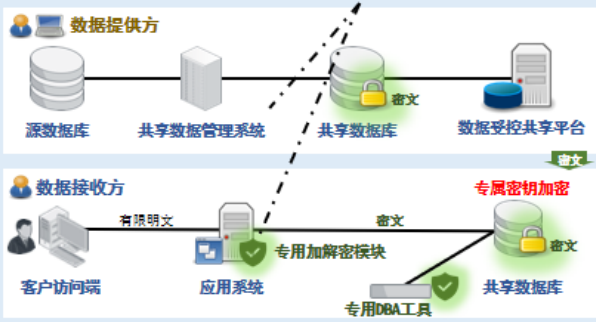
公开/删除
提供
传输
加工
使用
存储
收集



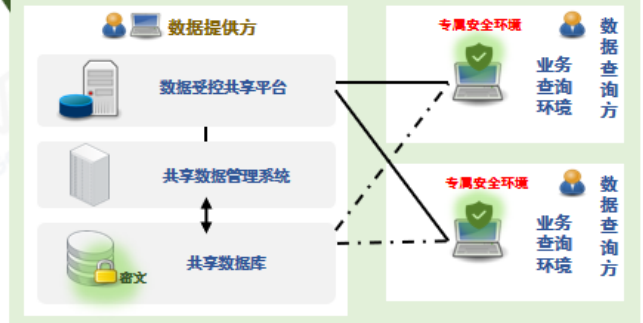
数据已成为企业间可以交易的商品，但交易后存在“数据失控”问题困扰着数据提供方：

- 数据的接收方有没有对数据进行滥用？
- 对方有没有再次将数据外发给第三方？
- 这些数据在接收方的有效期如何控制？

1. 离线场景：实现数据可见不可转



2. 在线场景：实现数据可查不可转



I. 识别　P. 防护　D. 检测　R. 响应　R. 恢复　C. 反制

跨域管控：部署模式不改变原有数据共享流程

炼石 CipherGateway

数据提供方

基础数据共享数据库 | 个人信息基础数据库 | 生产运营共享数据库 | 数据中台集中数据库 | 管理支撑共享数据库

数据中台-共享数据管理系统

来自内部的数据共享请求

数据中台发送数据给受控共享平台

跨域数据管控平台

关键技术

- 离线场景：实现数据可见不可转
- 在线场景：实现数据可用不可转
- 数据密态胶囊封装技术
- 受控数据共享专用终端技术
- 访问控制技术
- 数据水印泄漏溯源技术

核心功能

用户层	数据生产方	数据使用方	数据控制方	数据服务方		
安全层	数据加密	访问控制	权限管理	数字水印	数据脱敏/匿名化	数据审计
应用层	数据目录	数据溯源	数据授权	用量计费	交换通道	日志记录
支撑层	TEE 可信执行环境	同态加密 ME	数据信息管理	交换合约管理		
基础层	CA 服务	身份管理	密钥管理	区块链	交换通道	

来自外部的数据共享请求

数据胶囊及密钥下发

数据接收方

监管单位	专用加解密模块 应用系统 可信执行环境	专用DBA工具	数据使用
分子公司	专用加解密模块 应用系统 可信执行环境	专用DBA工具	数据使用
生态伙伴	专用加解密模块 应用系统 可信执行环境	专用DBA工具	数据使用

案例-政府高并发海量数据：某省近亿人民的防疫健康码数据保护

炼石 CipherGateway

项目背景介绍

客户行业：政府

项目名称：某省XX健康通数据保护

项目需求：

XX健康通服务近亿常住人口，从多渠道广泛收集大量个人隐私信息用于疫情分析，海量隐私数据在“裸奔”，需要针对系统建设的安全性进行等级保护测评和商用密码应用安全性评估。

项目建设内容：

简化复杂数据处理场景，保护数据中台核心数据，把控数据出入口，采用AOE插件实现实时流转数据入库加密、出库解密，且不影响对外接口和日常数据处理；采用SDK对SQL导入导出文件进行加解密处理，支持海量数据的实时处理。

涉及某省亿级健康码用户在40个应用服务中400多个字段的、200亿条数据加解密实施。

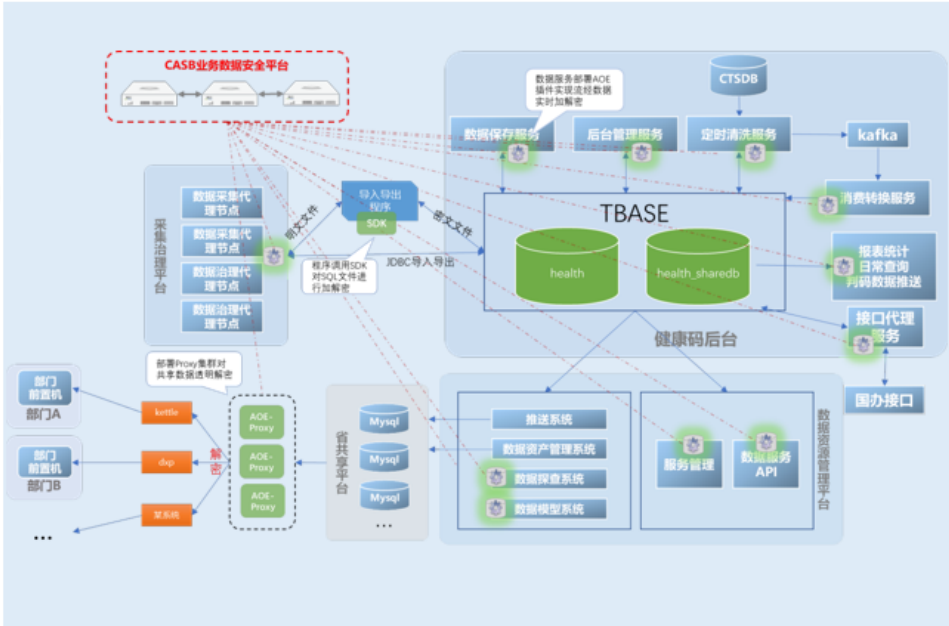
项目收益：

1、不影响数据中台业务的同时，保护了总数超1亿用户的防疫隐私数据；

2、支持丰富的数据处理场景，保障数据存储、处理、流转安全，支撑高峰时期5万多人同时使用。

3、满足国家关于政策合规、商用密码信息系统密码应用规范要求。

部署和实施方案



案例-金融核心交易：某商业银行重要数据保护

项目背景介绍

客户行业：银行

项目名称：某银行数据加密项目

项目需求：

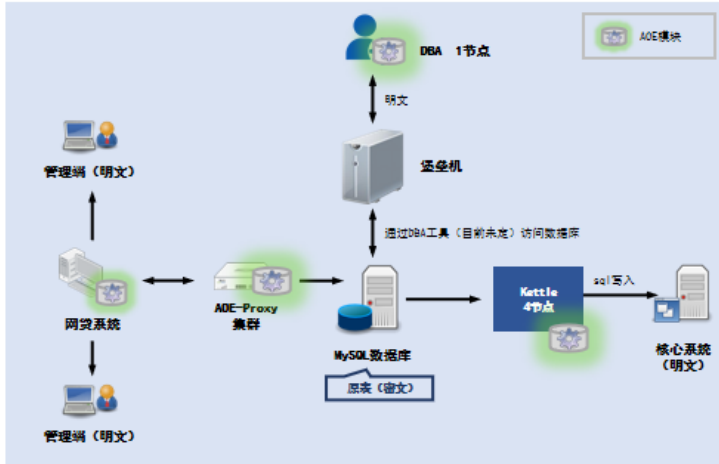
某商业银行掌握着海量用户个人隐私数据，但面对愈加严峻的内外部数据安全威胁，银行亟需兼顾用户隐私保护与经济发展需求之间的平衡，在运用数据为用户提供产品服务的同时，强化个人敏感信息的保护。

项目实施痛点：

- 待保护的数据量达到亿级，且数据存在形式多、访问人员众多、存储分散、易传播；
- 开发改造应用叠加数据安全，周期长、难度大、风险高；
- 信息系统环境复杂，涉及字段类型多、数据库种类多，涉及分库分表、存储过程、模糊查询等使用场景；
- 加解密过程不可影响或中断正常业务的运行。

项目建设方案

面向重要数据与个人信息保护，依托自主研发的AOE加解密模块，打造免开发改造应用的数据保护、高性能密码产品，构建高覆盖率的安全增强点组合，有效保护结构化与非结构化数据，保障数据全生命周期安全，实现主体到应用内用户、客体到字段级的防护，打造实战化数据安全防护体系。



案例-运营商BOSS应用：某省移动BOSS数据保护

项目背景介绍

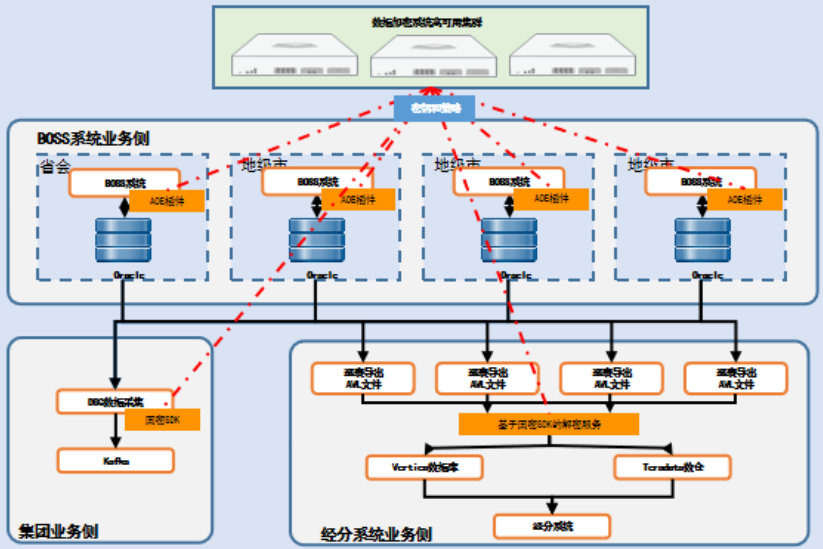
客户行业：运营商
项目名称：BOSS个人信息保护

项目需求：
针对BOSS客户资料表进行加密保护，在实现数据安全的同时，也符合两部委对于商用密码应用的考核要求。该方案已正式通过集团基线评审，获得集团大力支持。

项目实施痛点：
BOSS客户资料表加密，影响所有访问资料表的业务，涉及三方面协同改造（改造范围）：集团全网客户中心、省内CRM/BOSS/经分系统。

解决方案：
BOSS系统：1、应用安装AOE插件，应用写入数据实时加密，应用读取数据实时解密；2、历史数据可通过AOE客户端工具进行全量加密
集团：在省端的数据采集程序集成SDK，调用SDK解密接口，实现采集数据解密
经分： BOSS侧导出的密文AVL文件解密成明文文件，按照原逻辑解析明文文件，把数据写入vertica数据库和TD数据库

项目实施部署方案



案例-工业关键应用：“某骨干钢铁企业”工业数据保护

项目背景介绍

客户行业：钢铁制造

项目名称：工业制造系统数据安全保护项目

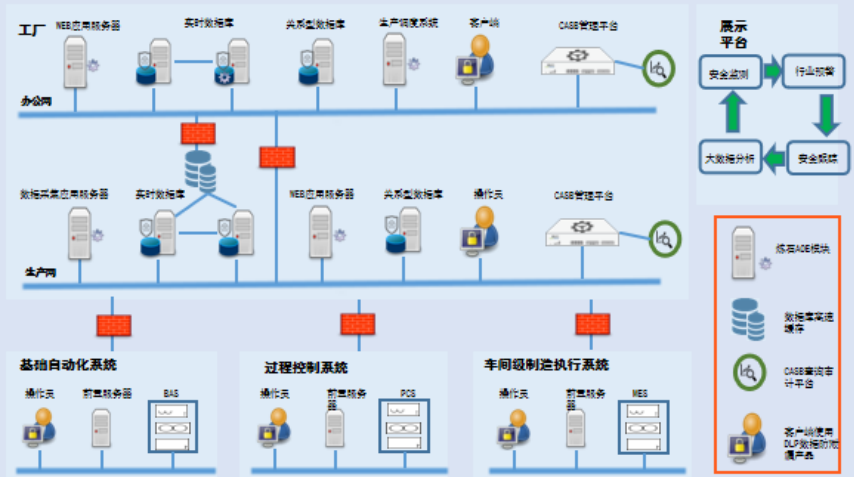
项目需求：

- 工业控制数据保护：生产过程中涉及的控制信息、工况状态、工艺参数等数据保护
- 研发数据保护：企业组织生产系统中涉及的研发设计、开发测试等数据保护

建设内容：

- 数据加密与访问控制：对制造配方敏感数据进行加密存储，确保数据在传输和存储过程中的机密性。对重要设备和系统包括BAS系统、PCS系统、MES系统实施细粒度访问控制，防止未经授权的访问。
- 数据审计：对所有工业等重要数据访问活动进行实时监控和记录，确保数据的完整性和安全性。实施日志审计机制，对数据访问日志进行收集、分析和存储。定期进行安全审计，检查数据安全策略的执行情况，发现潜在的安全风险和漏洞。对异常行为及时报警，以便快速响应和处理。

项目建设方案



案例-商业敏感数据应用：某知名酒店集团数据保护

炼石 CipherGateway

项目背景介绍

客户行业：文旅

项目名称：某知名酒店集团重要数据保护

项目需求：

某酒店集团内部信息系统众多，内含大量敏感个人信息及重要业务数据，对于集团来说，这些数据是宝贵的数据资产，有必要保护这些数据不被非法窃取、篡改和利用，因此，结合国家相关法规政策，亟需采用密码技术对应用系统中敏感数据进行加密保护。

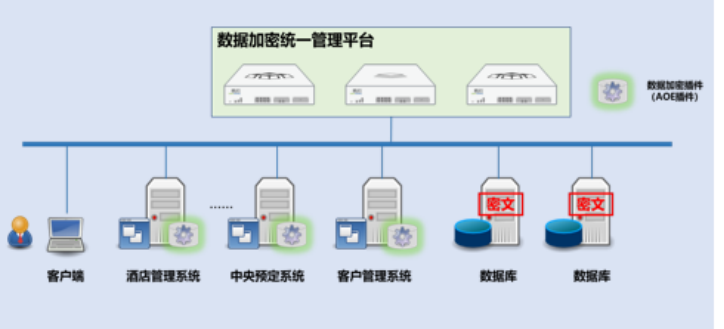
项目实施痛点：

- 集团前期采用了加密机路线，需要在应用系统端集成加密机的客户端SDK，严重影响了业务系统的运行效率；
- 业务系统纵横交错，数据存储散乱，加解密需要可配置、可视化的集中管控；
- 涉及敏感信息的信息系统多；信息系统分布于虚拟化、云化、实体机等各种环境中；涉及数据库种类多，如关系型数据库Oracle、MySQL和非关系型数据库MongoDB。待保护的数据量达到亿级，需要同时对业务实时数据及已有存量数据进行加密保护；
- 部分业务系统具有很强的时效性，面临低延时和高并发的挑战，要满足加解密功能对于业务操作延时不超过10%。

项目建设方案

炼石提供CASB业务数据加密平台（包含数据安全插件（AOE插件）、数据安全管理平台、密钥管理系统），为集团提供快速敏捷且有效的数据安全保护方案，实现“分布式加密，集中式管控”。

在酒店众多应用系统的服务节点上分别部署AOE插件，旁路部署统一数据安全管理平台和密钥管理系统。数据安全管理平台为插件执行点统一下发加解密及脱敏策略；密钥管理系统负责统一管理加解密所需的密钥，实现密钥全生命周期的管理，保障密钥的安全备份与恢复，从而完成集中的数据加密管理平台的建设，实现应用系统敏感数据的分布式加密/脱敏，安全策略的集中式管控。



案例-医药制造应用：某药企商业秘密数据保护

项目背景介绍

客户行业：医药制造
项目名称：创新药研制数据安全保护项目

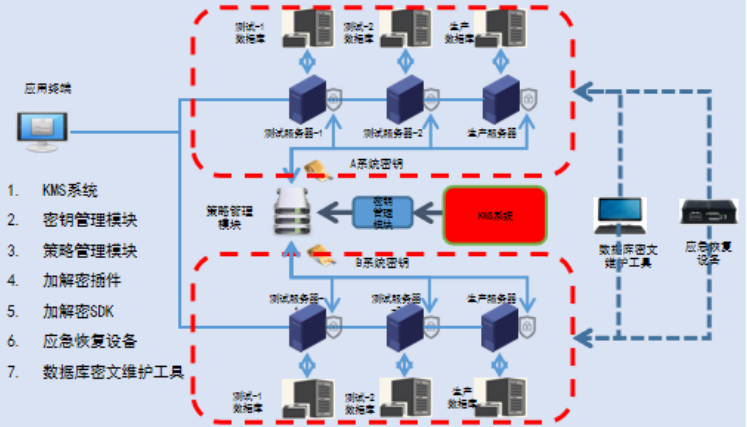
项目需求：

- **制药行业合规的要求** ：药企的信息化和数据可靠性的关系
  国际制药行业领先企业的RUTH数据可靠性管理策略
  GMP合规：国际药品生产质量管理规范
  GSP合规：药品经营质量管理规范

- **创新药研究的安全需求：**实验数据、分析数据的线上流转保护
  创新药物研究的数据安全保障

- **一体化工厂协同效率提升的安全需求：**工业4.0下大数据支持的 计产销联动
  经营信息数据安全、药品溯源保障
  打破数据孤岛，加强数据流转：数据的流动转移的安全保障
  跨领域协同联动：IT---OT，边界模糊情况下数据安全保障

项目建设方案



公司定位：炼石就是数据安全

炼石是以**“免改造”为特色的数据安全产品厂商，国家级专精特新“小巨人”，**自研灵活投送多重安全能力的免改造平台，帮政企客户打造领先数据安全保护体系，敏捷交付密评密改合规。

2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025
公司成立 安天投资	应用融合 数据安全	Pre-A轮 国科嘉和投资	免改造 数据安全	免改造 数据保护平台	A轮 腾讯投资	数据安全 主平台	A+轮 领航新界投资	数据安全 产品+服务	全域 数据保护	下一代 密服平台
威胁检测 系统安全 强化炼石系统安全能力		强化炼石安全算法能力		腾讯云 腾讯安全 产品共研深度融合			LHVENTURES 领航新界 拓展市场战略布局		朗玛峰大学 运营赋能增效发展	
安天 网络安全民企国家队		国科嘉和 中科院直投基金		Tencent 腾讯 产业互联网巨头			重庆科技成果转化基金 国资		朗玛峰创投 科创赋能专业基金	

炼石
CipherGateway

炼石免改造数据安全

www.ciphergateway.com

010-88459460

sales@ciphergateway.com