AI 管理体系

# ISO 42001 认证
# 概述指南

# 内容

# 探索 AI 环境

## 了解重塑技术、风险和责任的趋势

**人工智能（AI）**越来越多地应用于利用信息技术的所有部门，并有望成为主要的经济驱动力之一。这一趋势的结果是，某些应用程序可能会在未来几年引发社会挑战。从金融和制造到医疗保健和物流，AI 正在推动自动化、效率和决策方面的进步。生成式 AI 和机器学习的广泛采用开启了新的可能性，但也加速了对更强大**治理**、更高**透明度**和更明确**问责制**的需求。

### 全球 AI 格局正受到三个融合趋势的影响：

#### 加速实施

企业正在快速将人工智能嵌入核心运营。据 IBM 称，42% 的企业已在探索或积极部署生成式人工智能。然而，这种步伐往往超过了正式治理结构的发展速度，从而导致监督、质量保证和风险管理方面的缺口。

#### 不断发展的法规

各国政府和监管机构正在通过新的框架来应对日益增长的社会担忧，以确保人工智能系统的安全、公平和可解释性。2024年通过的《欧盟人工智能法案》为基于风险的监管开创了先例，全球范围内也正在推行类似的举措。传递的信息很明确：对人工智能的信任必须赢得，而不是被预设。

#### 日益严格的审查和道德期望

从数据隐私和知识产权到偏见和问责，各组织都面临着巨大的压力，需要证明其对人工智能的使用符合道德规范、负责任且安全。公众信任脆弱不堪—任何失误都可能迅速导致声誉受损、监管制裁和机遇错失。

这些趋势标志着一个转折点。AI 已从试点计划转向战略基础设施。随着这种转变，需要结构化的、体系范围的治理，这种治理可以雄心勃勃地扩展，并且经得起审查。

# 什么是 ISO 42001?

## 首个人工智能国际管理体系标准

ISO 42001 是世界上第一个针对人工智能的管理体系标准，提供了一个结构化的框架，帮助组织负责任地管理人工智能。

该标准专为开发、部署或依赖 AI 的组织而设计，规定了建立、实施、维护和持续改进人工智能管理体系（AIMS）的要求。

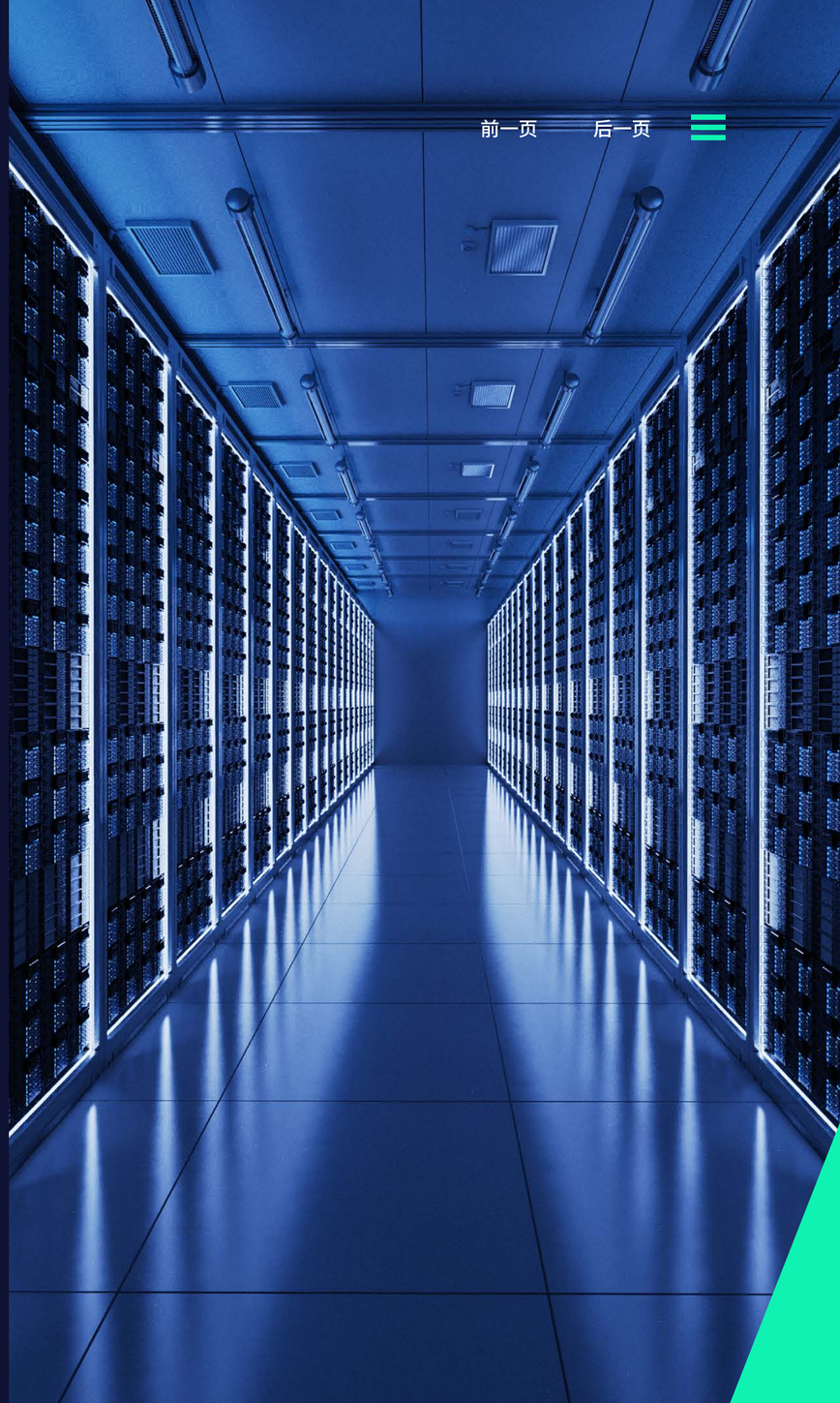
无论您是将 AI 集成到业务运营中，还是提供支持 AI 的产品和服务，ISO 42001 都可以帮助您嵌入道德原则、管理风险，并符合全球对可信 AI 的期望。

该标准的关键要素包括：

- **治理和问责制**–定义 AI 相关活动的角色、责任和监督
- **基于风险的控制**–识别和解决整个 AI 系统生命周期中的风险
- **透明度和可解释性**–支持清晰传达 AI 系统如何运作和做出决策
- **持续改进**–随着时间的推移，使用反馈、监控和绩效数据来加强您的 AIMS

ISO 42001 遵循附录 SL 结构，使其易于与 ISO 9001、ISO 27001 或 ISO 45001 等其他标准整合，从而为整个组织的风险提供一致和协调的方法。

通过采用 ISO 42001，组织可以展示对负责任创新的明确承诺，在日益 AI 驱动的世界中与客户、合作伙伴和监管机构建立信任。



# 为什么要获得认证?

**ISO 42001 认证不仅仅是一种形式。**
这是一个独立的、全球公认的保证标志，表明您的组织正在负责任地、有效地管理人工智能。

**展示对最佳实践的承诺**
认证表明您认真构建符合道德、透明且治理良好的 AI 系统，符合国际期望。

**与客户建立信任和信誉**
对于许多行业来说，认证正在成为贸易许可证。它向客户、合作伙伴和利益相关者保证，您的 AI 实践是安全的、负责任的并且管理良好。

**支持隐私、道德和信息安全目标**
认证流程有助于将 AI 治理嵌入到更广泛的风险管理系统中，从而加强数据保护、负责任的创新和道德监督。

**面向未来的合规性**
随着全球 AI 法规的加速发展，ISO 42001 认证有助于您的组织领先于不断变化的法律和行业要求，从而降低风险并支持长期弹性。

# ISO 42001 关键要求：

## 附录 SL 条款结构

**ISO 附录 SL 结构由 10 个条款组成。管理体系标准（包括 ISO 42001）中的所有内容都必须满足所有十个条款的标准，才能遵循附录 SL 框架。条款分为：**

**第 1 条**
**范围**

定义 AI 管理体系的预期结果及其在组织内的适用性。

**第 2 条**
**规范性引用文件**

列出对 ISO 42001 应用至关重要的参考标准。

**第 3 条**
**术语和定义**

提供整个标准中使用的核心术语的定义，以确保共同理解。

**第 4 条**
**组织环境**

考虑内部和外部因素、利益相关方的期望以及 AI 系统的使用范围。

**第 5 条**
**领导作用**

概述最高管理层在政策、资源和促进负责任的 AI 文化方面的责任。

**第 6 条**
**策划**

解决组织如何识别和应对 AI 风险和机遇，并设定 AI 目标。

**第 7 条**
**支持**

涵盖有效治理所需的资源、能力、意识、沟通和文档。

**第 8 条**
**运行**

定义如何实施与 AI 相关的控制措施，包括风险评估和系统影响评估。

**第 9 条**
**绩效评价**

需要监控、测量、内部审核和管理审查以评估体系绩效。

**第 10 条**
**改进**

详细说明组织对持续改进、不合格项处理和纠正措施的方法。

**在大多数情况下，这些条款使用相同的核心文本，无论它们适用于何种标准，并且共享通用术语和定义，以促进管理体系标准之间的一致性和兼容性。对于 ISO 42001，这确保了 AI 治理嵌入到熟悉且经过验证的管理体系结构中。**

# 在 AI 背景下满足 ISO 42001 要求



与其他 ISO 管理体系标准一样，ISO 42001 围绕第 4 至 10 条建立，涵盖组织环境、领导作用、策划、支持和持续改进等领域。ISO 42001 的独特之处在于，这些熟悉的概念如何针对与人工智能相关的挑战和风险进行定制。

以下部分总结了这些条款在 AI 环境中的适用情况—从道德治理和数据透明度到风险管理和生命周期监督。

## 组织环境

ISO 42001 要求组织通过了解其内部和外部环境来定义其 AI 管理体系的范围。这包括法律义务、行业特定的 AI 风险、文化和道德规范、利益相关方的期望以及组织在 AI 生命周期中的角色—无论是作为开发人员、部署者还是用户。这些见解塑造了 AI 的治理方式，并有助于确定哪些控制措施是相关的。还鼓励组织评估更广泛的社会问题，例如气候变化或数字公平，是否应该为其方法提供信息。

## 领导作用和治理

ISO 42001 的一个核心要求是领导层对负责任地使用 AI 的可见和持续承诺。高级管理层必须制定明确的 AI 政策，设定与组织价值观一致的目标，并确保在整个 AI 生命周期中定义角色和职责。这包括保持对风险评估、影响评估和人工智能使用的监督，特别是在敏感或高风险环境中。高层的积极参与有助于确保将治理嵌入整个企业，而不是事后才考虑。

## 道德和透明度

道德考虑贯穿整个 ISO 42001。组织必须展示他们如何评估和解决 AI 对个人和社会的潜在影响。这包括考虑公平、非歧视和人类自主性，以及数据隐私和结果的透明度。必须采取控制措施来识别和减轻意外后果，并且必须定期记录、监测和更新这些做法的证据。

## 风险和机会管理

组织需要评估和处理 AI 特定的风险（包括影响合规性、声誉和安全的风险），同时还要确定创新和改进的机会。ISO 42001 认识到风险偏好和定义可能因行业而异，因此该框架具有适应性。重要的是，决策是有意识地做出的，以政策为指导，并随着时间的推移评估有效性。

## 持续改进

ISO 42001 遵循 Plan-Do-Check-Act（PDCA）模型。这意味着持续改进不是可有可无的，而是嵌入式的。组织必须监控其 AI 体系绩效，进行内审，并定期审查治理流程。无论是纠正措施还是适应监管变化，目标都是不断提高 AI 管理体系的适用性、充分性和有效性。

# ISO 42001 附录 A 控制措施

## 将负责任的 AI 转化为行动

除了 10 个主要条款外，ISO 42001 还包括附录 A 中的一组支持控制目标。这些控制措施旨在帮助组织实施实用、可审核的措施，以支持值得信赖的 AI 开发和使用。附录 A 控制措施分为 10 个类别，涵盖：

- **与 AI 相关的策略**–确保明确的领导意图和方向
- **内部组织**–定义治理角色和职责
- **AI 系统的资源**–管理 AI 中使用的工具、数据和基础设施
- **影响评估**–评估对个人和社会的风险
- **AI系统生命周期**–使系统设计与道德和监管目标保持一致
- **AI 系统数据**–确保质量、来源和相关性
- **为相关方提供信息**–促进透明度和问责制
- **AI 系统使用**–管理范围、意图和保护措施
- **第三方关系**–为供应商和合作伙伴设定期望
- **文档和可追溯性**–支持可解释性和可审核性

这些控制措施为实施 ISO 42001 提供了实际基础。它们可以适应不同级别的 AI 成熟度和复杂性，并构成就绪性评估和认证的关键部分。

# 实施 ISO 42001

## 实施 AIMS 的重点领域

ISO 42001 提供了一个框架，用于使用计划-执行-检查-行动 (PDCA) 模型实施负责任和有效的人工智能管理实践，该模型是所有基于附录 SL 的 ISO 标准共享的核心基础。

该标准通过互连的管理流程支持人工智能管理体系 (AIMS) 的开发，包括：

- **意识**

  组织应通过培训团队（从开发人员到决策者）了解 AI 的好处、风险和道德考虑因素，从而建立意识。
- **责任**

  必须分配明确定义的角色和职责，确保 AI 治理得到理解并嵌入整个系统生命周期 。
- **响应**

  应采取及时和协调的行动来管理风险，响应 AI 系统的影响，并在需要时采取纠正措施。
- **风险评估**

  必须以结构化、循证的方式评估与 AI 系统相关的风险，包括技术故障、意外结果或滥用。
- **系统设计与开发**

  AI 系统的开发和部署必须符合成文的政策和流程，支持合乎道德的使用、透明度和生命周期问责制。
- **治理和控制**

  组织应采用整体方法来管理 AI，包括数据质量、可解释性、公平性、可靠性和人工监督。
- **重新评估和持续改进**

  作为 PDCA 模型的一部分，应定期监控和审查性能。内部审计和管理审查有助于确保 AIMS 继续实现目标并适应新出现的风险和法规。

# 使用 ISO 42001 构建您的路线图

## 为有效的 AI 治理奠定基础

**实施 ISO 42001 始于明确的目标和强大的领导支持。这些早期步骤有助于确保您的 AIMS 既实用又与组织的目标保持一致。**

### 1. 获得领导承诺

高级管理层必须领导 AIMS—设定明确的目标，定义组织的成功标准。无论是加强监督、满足监管期望还是建立信任，领导层都应该分配所需的资源和方向，将负责任的 AI 使用嵌入到整个企业中。

### 2. 了解您的 AI 环境

评估 AI 在整个组织中的开发、部署或使用方式。考虑法律义务、行业特定风险和利益相关方的期望。这种理解应该为您的资源提供信息，包括负责任和有效地管理 AI 所需的时间、技能和预算。

### 3. 评估培训要求

AI 治理涵盖技术、伦理道德、法律和运营领域。让跨职能团队参与进来，并确保提供量身定制的培训-从针对更广泛团队的认知到针对专家的审核培训。跨角色构建功能是有效、可扩展实施的关键。

### 4. 定义 AIMS 范围

明确您的 AIMS 将涵盖哪些团队、系统和地理位置，包括内部开发的 AI、第三方工具和高风险用例。明确定义的范围将使治理保持重点并适合目的。

### 5. 规划和实施您的 AIMS

制定一个切合实际的实施计划，涵盖时间表、职责和所需资源。引入必要的控制、文档和治理流程。构建定期审查、反馈循环和绩效监控，以确保您的 AIMS 随着您的业务和 AI 环境的发展而发展。

### 6. 进行差距分析

审查您当前的治理、风险和合规性框架，以确定您的组织在哪些方面已经达到标准，以及哪些方面存在差距或漏洞。这种洞察力将有助于确定改进的优先级并避免重复。

### 7. 预约认证审核

一旦您的 AIMS 就位，请安排 LRQA 的 ISO 42001 认证审核。我们的两阶段流程将评估您的体系设计和实施，帮助您向利益相关方展示问责制、诚信和负责任的创新。

### 8. 嵌入持续改进和响应

成熟的 AIMS 不仅仅包括初始实施，它还需要持续学习和改进的机制。建立反馈循环、定期审核和绩效指标，以适应不断发展的 AI 风险和技术。实施特定于 AI 的事件响应计划，以确保错误、偏差或系统故障得到快速、透明和有效的解决。

# 将 ISO 42001 与现有管理体系相结合

在已经有效的基础上进行构建。
加强 AI 和信息安全方面的治理。

对于许多组织来说，ISO 42001 不是一个起点，而是一个自然的延伸。如果您已经持有 ISO 27001 认证，那么您将处于高效整合 ISO 42001 的有利位置。

ISO 27001 为管理信息安全风险提供了一个经过验证的框架，其对数据保密性、完整性和可用性的控制直接支持 ISO 42001 中的许多要求。通过结合这两个标准，您可以创建一种统一的治理方法，同时解决您的信息和 AI 相关风险。

## 为什么集成有意义

### 共享结构

ISO 42001 遵循附录 SL 架构，与 ISO 27001、ISO 9001、ISO 45001 等中使用的高级架构相同。这实现了政策、领导、规划、运营和评估的一致性。

### 高效利用资源

集成有助于减少审核、文档和内部审查之间的重复。团队可以协调报告、目标和控制，从而简化合规性并改善监督。

### 一致的风险管理

这两个标准都要求采用基于风险的方法。ISO 27001侧重于信息资产，而 ISO 42001 则将其扩展到 AI 系统，包括如何使用数据来训练、验证和运作它们。

### 更强的系统弹性

集成方法可以更容易地发现系统性问题，全面解决这些问题，并向客户、监管机构和合作伙伴展示联合治理模型。

# LRQA 劳盛 ISO 42001 服务



## 培训

通过适配不同经验水平的多样化课程体系，以多元教学方式助您构建ISO 42001专业知识。

## 差距分析

认证前预审服务（可选）：由资深审核专家协助识别体系中的关键风险项与薄弱环节。

## 认证

我们根据ISO 42001的要求审核您的AIMS，特别关注附录A控制措施及其对您的管理体系的影响。

## 整合管理体系

如果您实施了多个管理体系，则可以从整合的审核和监督计划中受益，这是一种更高效、更具成本效益的风险管理方式。

**查看可用课程**

# 为什么选择LRQA?



在 LRQA，我们帮助组织制定稳健的、面向未来的风险管理计划，以实现安全、负责任和有效的人工智能和技术。

从确保遵守不断发展的 AI 法规到加强数据安全和将最佳实践嵌入治理，我们提供自信地推动创新所需的保证，帮助企业集成 AI 和技术，同时主动管理风险。

## 实地专业知识

我们的解决方案由专门从事网络安全、合规性和供应链风险管理的全球专家团队提供，帮助您应对 AI 相关风险，满足不断变化的监管要求，并将负责任的 AI 实践整合到您的运营中。

## 基于解决方案的合作伙伴关

我们不仅提供认证，还与您并肩合作，将 AI 治理整合到更广泛的风险和合规战略中。我们量身定制的方法确保 AI 和技术有助于推动可持续增长，同时满足不断变化的法律和道德期望。

## 持续保障

AI 驱动型风险需要持续监督。我们的实时风险管理方法可以主动解决问题，减少业务中断并增强弹性。我们互联的风险管理解决方案组合可帮助企业超越监管要求，将 AI 风险管理嵌入日常运营中。

## 数据驱动的决策

我们利用数字平台和分析来更深入地了解整个企业的 AI 风险。我们的人类智能通过数据驱动型工具得到增强，可帮助组织识别漏洞、预测未来风险并自信地做出明智的决策。

## 关于LRQA：

LRQA 是全球领先的保障业务合作伙伴，数十年来在审核、检验、咨询和网络安全服务方面积累了极为丰富的专业知识。凭借基于数据的洞察力，我们致力于帮助客户解决各种重大的业务挑战。

LRQA 的业务遍及全球150 多个国家，建立了超5000 人的团队。我们的合规、供应链、网络安全和ESG 专家屡获殊荣， 他们帮助全球61,000 多家客户预测、降低和管理风险，覆盖几乎所有行业。

我们的一举一动，都旨在为员工、客户、社会和地球创造更美好的未来。

## 联系我们

网站：www.lrqa.com.cn
邮箱：marketing.gc@lrqa.com
全国热线： 021-60233650 / 021-60233657
地址：上海市黄浦区南京西路288号创兴金融中心12楼

扫码关注官方微信