中国政法大学
法律与科技创新研究室
创新合规&人本关怀
Law360.ai

创新天使团
Meta360 DAO

@法律与科技创新研究室

LAW360

💕专注GRC（AI风险与法律合规）
为可持续创新大航海护航保平安💕

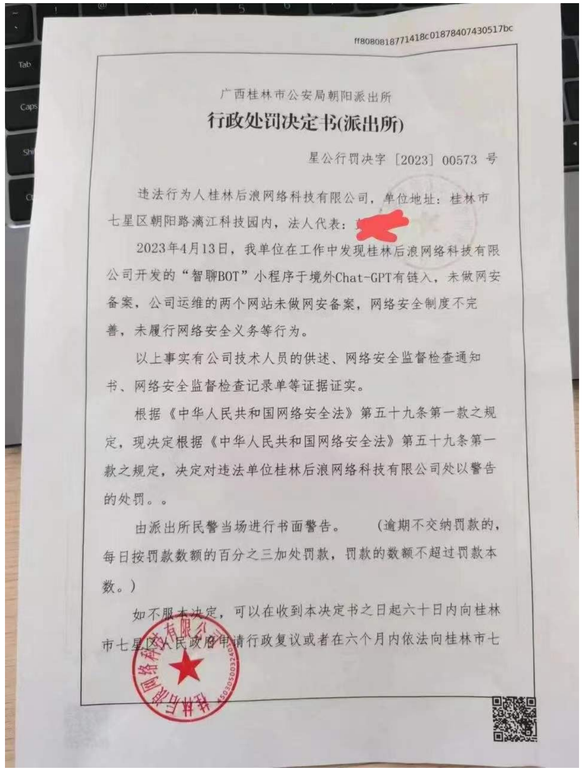
AI备案与评估解读 V1.0

- AI企业合规风险
- 《暂行办法》出台背景
- 重点法条解读
- 备案流程简述
- 人工智能合规评估框架

AI备案实战宝典专题群

扫码申请入群

AI监管开始实施，合规风险不容小觑



广西桂林市公安局朝阳派出所

行政处罚决定书(派出所)

星公行罚决字〔2023〕00573 号

违法行为人桂林后浪网络科技有限公司，单位地址：桂林市七星区朝阳路漓江科技园内，法人代表：

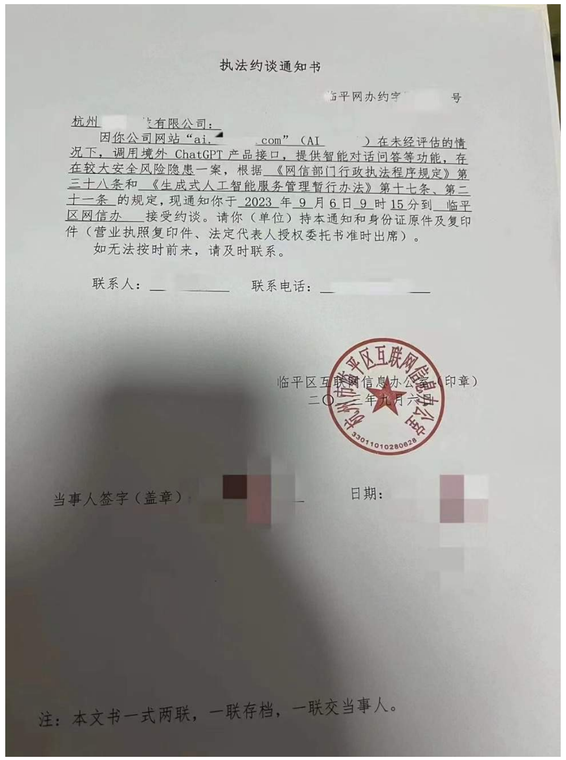
2023年4月13日，我单位在工作中发现桂林后浪网络科技有限公司开发的“智聊BOT”小程序于境外Chat-GPT有链入，未做网安备案，公司运维的两个网站未做网安备案，网络安全制度不完善，未履行网络安全义务等行为。

以上事实有公司技术人员的供述、网络安全监督检查通知书、网络安全监督检查记录单等证据证实。

根据《中华人民共和国网络安全法》第五十九条第一款之规定，现决定根据《中华人民共和国网络安全法》第五十九条第一款之规定，决定对违法单位桂林后浪网络科技有限公司处以警告的处罚。。

由派出所民警当场进行书面警告。（逾期不交纳罚款的，每日按罚款数额的百分之三加处罚款，罚款的数额不超过罚款本数。）

如不服本决定，可以在收到本决定书之日起六十日内向桂林市七星区人民政府申请行政复议或者在六个月内依法向桂林市七



执法约谈通知书

临平网办约字 号

杭州 有限公司：

因你公司网站“ai. com”（AI ）在未经评估的情况下，调用境外 ChatGPT 产品接口，提供智能对话问答等功能，存在较大安全风险隐患一案，根据《网信部门行政执法程序规定》第三十八条和《生成式人工智能服务管理暂行办法》第十七条、第二十一条的规定，现通知你于 2023 年 9 月 6 日 9 时 15 分到 临平区网信办 接受约谈。请你（单位）持本通知和身份证原件及复印件（营业执照复印件、法定代表人授权委托书准时出席）。

如无法按时前来，请及时联系。

联系人： 联系电话：

临平区互联网信息办公室（印章）

二〇二三年九月六日

当事人签字（盖章） 日期：

注：本文书一式两联，一联存档，一联交当事人。

公司名称	2023年4月	2023年5月	2023年6月	2023年7月	2023年8月	2023年9月	8月到9月
liblib		40	75	370	560	207	-63%
海艺AI		49	292	556	506	450	-11%
吐司		0	16	99	115	58	-50%
Vega	87	93	110	102	92	54	-41%
触手AI		2.2	12	21	23	17	-26%
无界AI	40	32	29	25	18	12	-33%
造梦日记	15	11	11	8	7.2	5	-31%

国产独立AI网站流量情况

- 国产图像AI网站遭受了严重打击。
- 可以理解为半归零状态。

《暂行办法》出台背景

- **整体思路：**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 》一脉相承
- **监管主体：**七部委共同监管
- **分类分级**：区分**技术**和**服务**、面向**公众**和与**非公众**
- **参考借鉴：**科技部《科技伦理审查办法》、欧盟《AI法案》、美国《AI风险管理框架》、英国《支持创新的AI监管方式》、意大利对OpenAI的隐私保护和信息留存限制措施等

相比征求意见稿：

- 适用范围减少、主体义务减轻、监管要求柔化

重点法条解读

服务规范（9~15）

- **第九条：**减轻服务提供者义务，内容生产者责任改为网络信息内容生产者责任，责任限于网安层面
- **第十一条：**细化对个人信息收集、输入信息留存的规定

监督检查和法律责任（16~21）

- 第十六条：新增各部门依据职责分级分类监管
- 第十七条：关于**评估（双新评估）**和**备案**，修改了“向国家网信部门申报安全评估”的规定，改为“按照有关规定开展安全评估”
- 第二十条：新增对不合规的境外服务提供者的处置办法（GPT套壳管理的法规依据）

附则（22~24）

- 第二十三条：关于行政许可，看相关主管部门要求，无统一的牌照要求;
- 外商投资 to C场景的要服从对外资准入的限制性规定

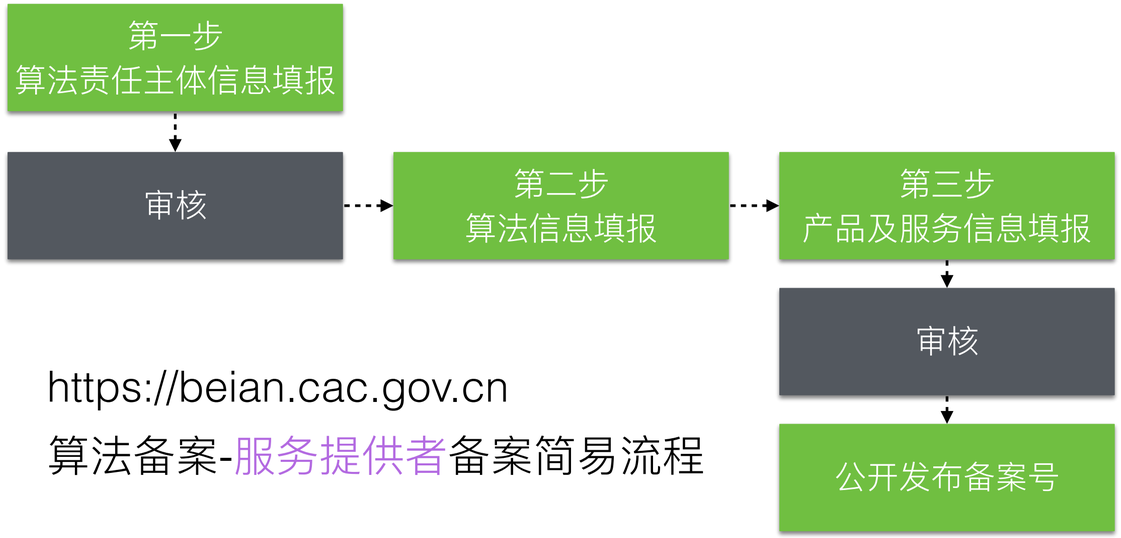
AI备案与评估

算法类型：生成合成类（深度合成）

备案主体：技术支持者 or 服务提供者

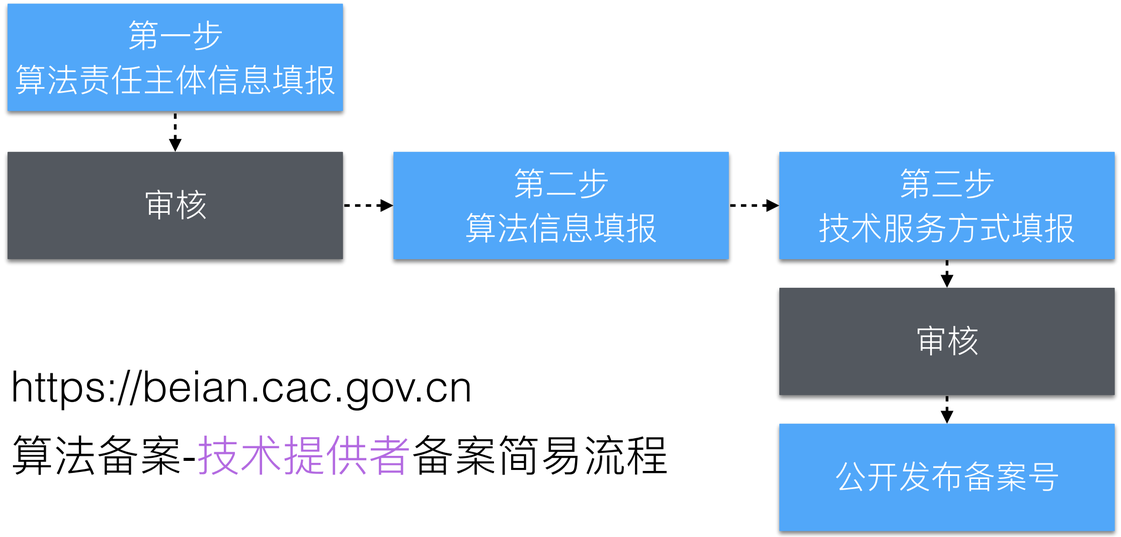
算法信息：

- 基础属性：类型、名称、应用领域等
- 详细属性：算法数据、算法模型、算法策略、算法风险与防范机制
- 产品及服务信息（服务提供者）：名称、服务形式、访问地址、服务状态、对象等
- 技术服务方式(技术支持者)：名称、访问方式、服务对象、服务频度等
- 其他附件：
  - 落实算法安全主体责任基本情况：算法安全专职机构以及制定的算法安全管理制度
  - 算法安全自评估报告：算法情况（算法流程、数据、模型和干预策略)、服务情况、风险研判、风险防控情况、安全评估结论等内容。
  - 拟公示内容：基本原理、目的意图和主要运行机制等



https://beian.cac.gov.cn

算法备案-服务提供者备案简易流程



https://beian.cac.gov.cn

算法备案-技术提供者备案简易流程

AI备案与评估

依据：《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》第三、五条

后续出台:《信息安全技术 生成式人工智能服务安全总体要求》《信息安全技术 生成式人工智能预训练和优化训练数据安全规范》《信息安全技术 生成式人工智能人工标注安全规范》，等国标形成配套指引

基础评估：

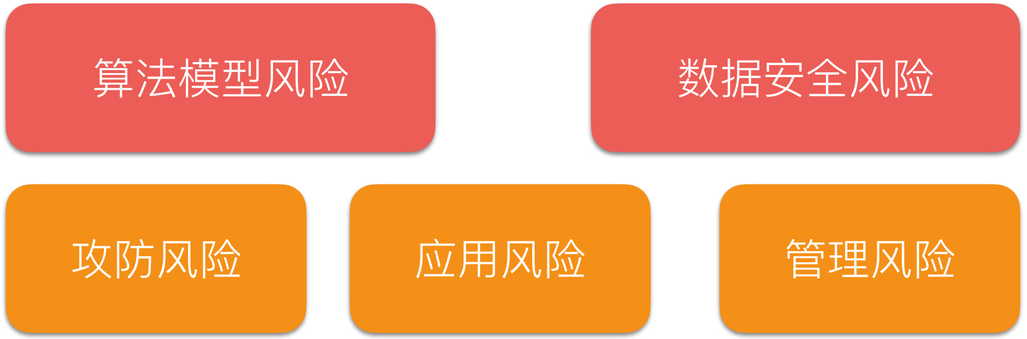
- 信息全生命周期管理能力
- 依据《互联网信息服务安全通用要求》

专项评估：

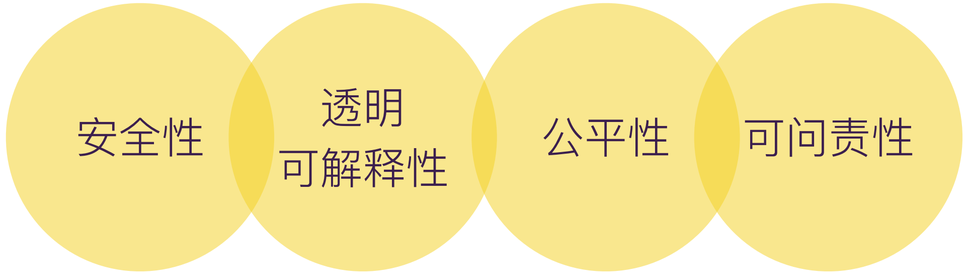
- 用户(信息源)账号管理：注册、使用、分级分类、追溯、违规处置
- 审核能力：审核管理(制度、人员)、审核效果(覆盖度、时效性重点内容)、综合识别技术(涉政有害、色情低俗、暴恐血腥)、专业识别技术(文本相似、关键词匹配)、内容审核、信息发布、监测预警、信息储存与销毁等
- 功能管理：发布、评论、即时通信、直播、推荐、搜索等各有规定适用
- 技术管理：管理制度、管理技术(内容安全技术、显著标记、合成鉴别技术等，描述针对模型训练数据和输入输出内容的审核过滤能力等)

人工智能合规评估框架

人工智能周期



人工智能风险



人工智能伦理准则

合规评估事项

制度

- 明确的目标 、利益相关者的参与 、用户沟通策略 、风险管理制度 、日志记录 、责任分工

组织

- 组织架构 人员配备 人员履职 教育与培训

数据

- 可追溯性、完整性、准确性、隐私合规、数据过度采集、数据偏差

性能

- 精确性、鲁棒性、公平性

安全

- 合规使用、攻击防范、开源框架漏洞、硬件安全、法律法规、标准规范

Law360.ai

中国政法大学
法律与科技创新研究室
合规创新&人本关怀

制度

明确的目标：是否从系统开发初期就有明确的目标，以确保实现预期结果。

利益相关者的参与：建立机制，在整个人工智能生命周期中，考虑来自利益相关者社区的不同视角，以降低风险，通过使外部利益相关者能够访问有关人工智能系统的设计、操作和限制的信息来提高透明度。

用户沟通策略：是否制定了沟通策略、向用户进行了政策解释、提供了退出选择权。有否建立机制，让使用者知悉人工智能系统所作决定的目的，准则及限制。

风险管理制度：实施专门的、持续的风险管理计划和应急响应计划，以系统地识别。分析和减轻风险。进行持续监测、验证和修正，评估人工智能系统可能被恶意使用、滥用或不当使用的风险。为第三者(例如供应商、最终用户、对象、分销商/供应商或工人)建立程序，以报告人工智能系统的潜在漏洞、风险或偏差。

日志记录：在整个人工智能生命周期中进行了日志记录

责任分工：明确定义人工智能系统涉及的人员角色、职责、分工,建立问责机制，以确保有效的操作，及时纠正和持续监督。

组织

组织架构：明确建立了人工智能安全风险的管理架构和监督架构

人员配备：根据责任分工为人工智能系统分配了明确的角色、职责和授权。

人员履职：人员是否按照定义的角色进行履职并落实责任，进行了必要的纠正和监督。

教育与培训：对组织人员和合作伙伴进行人工智能风险管理意识教育和培训，使其能够按照相关政策、程序和协议履行职责。

数据

可追溯性：保障数据的来源可追溯和可靠，评估第三方数据的评估和使用

完整性：评估数据偏差、采取措施以减少偏见，保障数据集的完整性等

准确性：采取必要的措施来保障数据与系统目标相称，及时更新，从多个数据集合并、提取、转换或进行其他相关操作时能够准确记录并正确执行

隐私合规：个人数据保护的做法和实践。是否制定了数据保护措施。

数据过度采集：是否确保收集了与系统目标相称的数据，并未扩大数据采集范围

数据偏差：评估系统运行中使用的所有数据的可靠性、质量和代表性，包括与人工智能系统数据相关的任何潜在的偏见、不平等和其他社会问题。

算法

可解释性：是否能解释AI如何实现预测。有没有向用户解释算法的运行和决策原理。

可追溯性：是否记录了过程(建模过程和算法部署)。你能否追溯哪个人工智能模型或规则导致了人工智能系统的决策或建议? 你可否追溯系统曾使用哪些资料作出某项决定或建议?

可重复性：你有没有制定验证及确认方法和文件(例如记录)，以评估和确保人工智能算法的可重复性，并能够重现结果?

性能

- **精确性：**是否满足预期需求。你是否制定了一个明确的过程来监控AI系统是否达到了预期的目标
- **鲁棒性：**模型是否稳健，是否保障在设计或技术故障、缺陷、停电、攻击、滥用、不适当或恶意使用等风险或威胁的情况下，人工智能系统不会产生对抗性、关键性或破坏性的影响(例如对人类或社会安全)
- **公平性：**在系统开发、应用等阶段是否制定了程序，以便于测试、检查、消除潜在的歧视和偏见。是否制定了一套策略或程序，以避免在人工智能系统中造成或加强不公平的偏见，无论是在输入数据的使用方面还是在算法设计方面?

安全

- **合规使用：**是否过度使用，是否过度放大了AI决策，忽略人的监督权和自主权是否强制使用，是否保障不存在用户不能拒绝使用AI系统的情况，是否恶意使用，是否保障系统不会被用于错误的目的，是否会被用于错误使用，是否保障系统不会用于错误的场合和群体，是否制定了程序测试、监控、处理和纠正AI系统对儿童、老年人、残疾人等群体造成的伤害人的参与程度是否合适
- **攻击防范：**各类攻击的防范，如对抗样本、逆向还原,是否评估过人工智能系统可能遭受的潜在攻击形式？是否考虑过不同类型的漏洞和潜在的攻击切入点？有没有采取措施确保人工智能系统在整个生命周期内的完整性、健壮性和整体安全性，以免受到潜在的攻击
- **开源框架漏洞：**对开源学习框架以及依赖库的安全性进行评估，和内部审计
- **硬件安全：**针对硬件攻击的预案和保障措施
- **法律法规：**确保人工智能系统符合相关的法律、法规
- **标准规范：**确保人工智能系统符合相关的标准、规范和指导方针

附录：人工智能合规指南（欧盟篇）

6 月 14 日星期三，欧洲议会通过了《欧盟人工智能法案》（AI Act）。就像 **GDPR** 引发全球数据隐私范式转变一样，《人工智能法案》将对组织开发、部署和维护人工智能系统的方式产生重大影响。

不应期望机器学习工程师了解该法规的全部细微差别。**这将是法律或合规团队的责任。**人工智能/机器学习从业者应专注于创新和推动新的机会，以提高效率和效果。然而，这些团队将承担新的职责，因此有必要对监管要求有基本的了解。

《人工智能法案》**并不直接针对模型本身**，因为监管不可能跟上人工智能研发和模型版本化的步伐。
相反，他们将寻求**规范组织开发和部署人工智能用例的人员和流程**。
这意味着机器学习和数据科学从业者的“日常”即将经历根本性转变，以确保人工智能用例得到正确记录、审查和监控。

AI用例分类分级

《人工智能法案》对特定用例进行了分类（不是 ML 模型）分为 4 个风险类别之一：不可接受（禁止）、高风险、中风险和低风险。

文档可访问性

新的文档系统将会出现，这些解决方案必须有效地弥合技术复杂性、业务级概念和需求之间的差距，最终在组织更广泛的背景下提供对技术工作及其局限性的全面理解。

AIGC的责任要求

欧盟议会在其人工智能法案版本中引入的最大变化之一是对部署基础模型和生成人工智能系统的组织提出了更明确的要求。

测试和人工评估

组织应开发自己的评估任务，并将其集成到测试套件中以确保模型质量。
此外，由非模型开发人员完成一组标准的评估任务可以实现不易被“操纵”的内部控制要求。

模型更新工作流程

任何用于“**高风险**”用例的人工智能都需要定义用于更新它的结构化流程。出于合规性目的，对机器学习模型的不同类型的更改可能会触发不同级别的审查。
虽然欧盟委员会已保证在《人工智能法案》颁布后将提供指导，但各组织应主动制定一份全面的“手册”，概述哪些类型的系统变更需要不同的重新评估工作流程和流程。

高风险用例的合格评定

对于内部评估，组织和服务提供商必须建立健全的质量管理体系，涵盖各种要素，例如全面的风险管理、认真的上市后监控、有效的事件报告程序（包括数据泄露和系统故障）以及识别风险的能力。
此外，应制定严格的数据管理测试和验证程序。在某些情况下，可能需要进行独立的第三方评估才能获得验证人工智能系统是否符合监管标准的认证——类似于医疗器械或食品等其他行业。

https://www.forbes.com/sites/forbeseq/2023/06/15/a-machine-learning-engineers-guide-to-the-ai-act

中国政法大学
法律与科技创新研究室
Law360.ai
创新合规&人本关怀

更多干货不要错过相关讲座

对话大咖

AI备案实战宝典

中国政法大学法学院法律与科技创新研究室

备案工作室

10/14 14:00 - 16:00

~~原价998元~~

限/时/价/格 1元

扫码申请入群

线上与线下活动同步，报名即可获得活动具体地址，近距离对话大咖

文天

AI产品已备案通过，曾就职网信办

现互联网大厂内容合规与战略负责人

分享主题

大模型备案实战漫谈

备案对AI公司的重要性，实践分享

赵渊

广联达科技股份有限公司法律事务总监

分享主题

AI在企业应用的法律挑战和应对策略

相关法律趋势解读，AI应用应对策略

贾广芳

政法大学法律与科技创新研究室研究员

人工智能与大数据合规业务部负责律师

分享主题

全景透视AI备案的点与面

四招教你梳理备案前提条件与撰写秘籍

主持人

朵姐

AIGC俱乐部发起人

微软认证AI讲师

互联网设计专家

入群福利

- 各大厂备案资料
- 算法备案制度范本
- 历史备案清单
- 备案法律法规大全

专题讲座名额有限扫码申请

法律与科技创新研究室

- 中国政法大学法学院“法律与科技创新研究室”“Research Office for Law and Technology Innovate of Law School of CUPL”简称“ROLTI”
- 由中国政法大学法学院联合中关村大数据产业联盟、清华 大学社会与智能治理研究院、北京市铭达律师事务所、武汉大学哲学院、 浙江大学法学院专家学者共同设立。

目标

- 致力于打造一个国际一流、国内领先的交叉学科研究平台和创新型孵化中心，
- “法律与科技创新研究室”促进法学专业与数智科学等相关专业的学术交流与融合，建设复合型团队、培养综合型人才，推动学术创新 与研究成果转化。

课题（部分）

- 人工智能技术引发的法律问题
- 大数据应用的法律问题
- 科技伦理的文明演变
- 法律职业智能辅助的技术与实践
- 学科交叉与实践融合的计算法学
- 计算法学作为一门新兴学科

法律与科技创新前沿

- 这些战略的目标都是在国家级策略的基础上提出的，涉及数 据、科技伦理、算法和法学领域。
- 联合强大的工业界、学术界取得合作共建。

主要发起单位

智库团队

- 法律与科技创新研究室，特聘专家顾问有焦洪昌教授、舒国滢教 授、陈波教授、熊明辉教授、王洪教授法学界、哲学界、逻辑学界资深 专家顾问

- **舒国滢教授** 中国政法大学学术委员会委员、法学院教授委员会 主席，国务院“政府特殊津贴”专家，主要研究法哲学、法理学
- **陈波教授**，武汉大学哲学学院人文社科讲席教授，博士生导师， 2018年8月，当选为国际哲学学院院士。 2021年10月，当选国际科 学哲学学院院士。
- **王洪教授**:中国政法大学法治文化专业(法治思维方向)博士生导 师，院学术委员会主席。北京市逻辑学会副会长。
- **熊明辉教授**，浙江大学光华法学院求是特聘教授、博士生导师， 教育部人文社会科学重点研究基地中山大学逻辑与认知研究所教授

中国政法大学 法学院
法律与科技创新研究室

相关服务

- 咨询服务
- 标准认证
- 培训服务
- 创投咨询
- 创新孵化

Law360.ai

中国政法大学
法律与科技创新研究室
合规创新&人本关怀

Law360.ai

中国政法大学
法律与科技创新研究室
合规创新&人本关怀

法律与科技创新研究室（Law360社区）

中国政法大学·法律与科技创新研究室和Meta360创新DAO 发起的跨学科智库，通过研究构建新一代法律产品和服务。让每个人都能受益于更加公平和人性化的法律服务

我们的目标

- 推广合规文化、普及法律知识
- 助推司法系统、法律服务、互联网平台、社区团体、政府机构、科技企业更好地帮助人们解决司法问题，保护人民合法权利
- 构建一个更加人性化、更加普惠、更加公平的法律环境

我们的理念和原则

- **以人为本设计理念：**致力打造更加人性化易于理解和使用的法律产品和服务
- **跨学科融合的创新：**通过跨领域合作，开发更加中和、创新的解决方案
- **知行合一创新探索：**通过创新孵化，将设计思维和高新技术应用于法律领域

复合型专家人才库

我们构建和研究新技术、服务和政策，为处理法律问题的人们提供帮助。

创新成果孵化

我们研究开发和测试更加人性化、简便的普惠法律服务、技术和新模式。

创新者联盟社区

我们协调能够扩展司法创新的机构、人员、想法和资源。为创新者提供支援

智库课题

研究、测试新技术（AI、区块链、元宇宙等）应用场景和创新案例
帮助创新团队设计更人性的产品和服务
推动透明可信和开源普惠AI创新

AGI前夜

AI创新研究（agi360.xyz）

打破专业知识和语言壁垒，让普通人也能理解和掌握AI前沿知识与技能，**帮助AIGC创新产品落地**

风险治理

GRC风险与合规管理（2L2B）

通过标准解读、合规指引，案例演示，培训，BaaS系统实施等方式降低创新企业风险管理与合规门槛

普惠服务

普惠法律创新研究（2G2C）

用更人性化设计和新技术普及法律知识，在线纠纷调解，区块链数字取证，法律援助降低司法系统门槛

安居乐业

EAP员工支援计划（人本关怀）

携手专业伙伴，以公益社区、创新产品等形式为企业和员工提供法律援助+心理咨询等公益性服务