

韩国《人工智能基本法》 经营者实务指南

合规义务、风险管理与商业机会

金京镇律师

AI书房 · kimkj.com · 2026

目录

1. 前言

第一部 经营者为什么现在必须读懂这部法律

2. 第1章 立法沿革与2026年修订：对经营者而言，有什么改变？从何时开始？

3. 第二章. 我是该法的适用对象吗：人工智能经营者的定义与适用范围

第二部 自我诊断：我们的服务适用哪些规制

4. 第3章. 第一阶段判定：我们是生成式人工智能经营者吗

5. 第四章. 第二阶段判定：我们是高影响人工智能经营者吗？

6. 第5章. 第3步判断：我们是高性能、最前沿人工智能经营者吗？

7. 第6章. 第4阶段判断：我们是需要指定韩国境内代理人的海外经营者吗

第三部 所有经营者的基本义务：确保透明性

8. 第7章. 事先告知义务：何时、何物、如何告知

9. 第8章. 生成物标识义务：水印、标签以及深度伪造的特例

10. 第九章 透明度指南实务、封闭网与开放网及例外事由

第四部 强化义务对象：高性能与高影响人工智能

11. 第10章. 高性能人工智能经营者的确保安全性义务

12. 第11章. 高影响人工智能经营者的五项责任

13. 第12章 高影响人工智能影响评估——非法律强制却为事实上的必须

14. 第13章. 确认是否属于高影响的申请程序

第五部 经营者面临的相邻法律风险

15. 第14章. 训练数据与著作权：合理使用条文、韩国首起诉讼、海外判例的警告

16. 第15章. 与《个人信息保护法》的连接点

17. 第16章 行业邻接规制的重复

第六部 经营者可以获得的支持与机会

18. 第17章. 学习用数据、基础设施、集聚园区，获得政府支持的方法

19. 第18章. 公共采购的优先考虑与公务员免责条款 —— 开启B2G市场的钥匙

20. 第19章. 弱势群体无障碍支持与费用支持，2026年修订开启的新市场

21. 第20章 创业·专业人才·创业投资母基金·人工智能研究所 —— 获取资金与人才的方法

第七部 违反时会发生什么：制裁与风险管理

22. 第21章. 事实调查的启动条件与资料提交应对

- 23. 第22章. 行政罚款、纠正命令、罚则体系与辅导期的现实
- 24. 第23章. 应对监督的文件管理——应当保管哪些文件，保管多少？

第八部 海外扩张与外国竞争者的读法

- 25. 第24章. 与EU AI Act并列看待的韩国法律
- 26. 第25章 2026年欧盟综合修正案对韩国经营者的实务启示
- 27. 第26章 美国科罗拉多州、犹他州法律与NIST AI RMF：全球经营者的准据点

第九部 行业别经营者实务指南

- 28. 第27章 金融人工智能经营者
- 29. 第28章. 医疗及生物人工智能经营者
- 30. 第29章 平台与生成式AI业务经营者
- 31. 第30章. 公共部门供应经营者

附录

- 32. 附录 A. 经营者自检流程图
- 33. 附录 B. 条文-施行令-指南对应表
- 34. 附录 C. 经营者类型分类清单
- 35. 附录 D. 施行日程时间线
- 36. 附录 E. 风险管理方案及影响评估文件格式样本

前言

2026年1月22日，韩国《人工智能基本法》正式施行。同年7月21日，修改后的法律也将生效。无论是开发还是使用人工智能的经营者，都应当了解这部法律为自己课以了何种义务、带来了何种机遇。

本书由律师面向经营者撰写，是一本侧重实操的实务指南，而非偏向理论的法学教科书。书中回答了诸如「我们的服务是否属于高影响人工智能？」、「如何进行事先告知？」、「如果进入欧盟市场，还需要额外准备什么？」等具体问题。

许多问题仅靠阅读法律条文很难找到答案。必须将施行令、公告、指南以及尚未敲定的后续立法结合起来看，才能看清全貌。书中还对比了《EU AI Act》与中国的监管规定。因为在全球市场开展业务的韩国企业，仅了解韩国法律是远远不够的。

本书力求通俗易懂。在无法避免使用法律术语的地方，均在旁附上了释义。书中完整记录了条文编号，想要核对原文的读者可以直接在国家法令信息中心检索相关条款。

这部法律目前处于施行初期。过渡期将运行一年以上，施行令与公告也在陆续完善。本书中的部分解读与展望可能会随着后续立法或有权解释的变化而有所调整。尽管存在这些局限，但对于经营者当下应该着手准备的事项，书中仍给出了明确的判断。

希望本书能为人工智能经营者降低法律风险，并在争取政府支持与发掘新市场机遇时提供切实的帮助。

2026年7月

金京镇

第一部 经营者为什么现在必须读懂这部法律

第1章 立法沿革与2026年修订：对经营者而言，有什么改变？从何时开始？

2024年12月26日下午，国会全体会议厅的大屏幕上显示了电子投票结果。在席264人，赞成260人，反对1人，弃权3人。《关于促进人工智能发展及营造信任基础等的基本法》在国会获得通过。继欧盟《人工智能法案》（EU AI Act）之后，世界上第二部全面规制人工智能的法律由此诞生。这场表决历经了四年半的时间。在第21届国会搁浅，在第22届国会将19项法案合而为一，在科学技术信息广播通信委员会（以下简称“科广委”）审议室里，朝野双方为一个个术语针锋相对。甚至在法律通过后、距离施行仅剩两天时，又进行了一次修订。

1. 四年半的历程：从2020年提案到2024年通过

2020年7月，第21届国会首次提出了人工智能相关的基本法案。2021年7月，郑必模等23名议员提出了《关于培育人工智能及营造信任基础等的法律草案》，随后多名议员也陆续提出了宗旨相似的法案。问题在于方向。主张培育产业的一方与主张优先建立安全规制的一方未能找到契合点。朝野之间关于规制水平的分歧历经三年多仍未弥合，随着2024年5月第21届国会任期届满，相关法案全部自动废止。

第22届国会在开院次日便行动起来。2024年5月31日，安哲秀议员提出了《关于培育人工智能产业及确保信任的法律草案》（议案号：2200053）。执政党国民力量党方面，郑点植、金成愿、裴俊英、郑熙溶等议员提出了5项；在野党共同民主党等方面，赵仁哲、闵馨培、权七胜、李海旻、崔敏姬等议员提出了12项；朝野共同提出了2项。截至2024年11月21日，共提交了19项法案。科广委下属的信息通信广播法案审查小委员会将这19项法案整合为一个委员会替代法案，经11月26日科广委全体会议审议后，于12月17日在法制司法委员会获修正通过。12月26日提交至全体会议，以压倒性赞成票获得通过。

2. 科广委审议室的三项决定

在将19项法案融为一体的过程中，有三个争议点十分激烈。

韩国看重“影响”，而欧盟看重“风险”。EU AI Act使用了“high-risk（高风险）”一词。韩国法案在初期也曾出现过“高风险”，但有人担忧这会给技术本身打上负面烙印。李海旻议员发言称：“全球趋势之所以走向‘影响’，是因为影响既能包含危险性，也能涵盖范围。”最终方案采纳了“高影响人工智能”这一概念。这意味着焦点将放在社会影响力上，而非技术本身。

是事前防范，还是事后追责？初期法案中包含强制要求进行事前影响评估的条款。企业方面反驳称这会削弱全球竞争力。随后达成了妥协方案。将高影响人工智能的影响评估设定为“自主努力义务”，但构建了在公共采购中对履行了影响评估的经营者予以优先考虑的机制。这是一种“软性规制”，虽然法律未作强制，但若想将高影响人工智能产品供应给公共部门，实际上必须接受评估。

经过一番争议，行政罚款的标准被定得较低。EU AI Act的最高行政罚金为3500万欧元或全球营业额的7%。而韩国《人工智能基本法》的最高行政罚款仅为3000万韩元。其背后原因是认为韩国国内AI产业尚处于初期成长阶段，过度的处罚会萎缩创新。至于3000万韩元这个金额能否给大企业提供遵守的动力，这种疑问在法律施行之后依然不断出现。

3. 2025年，制定施行令和指南

2025年1月21日，韩国《人工智能基本法》作为第20676号法律予以公布。施行日期为一年后的2026年1月22日。这期间的一年被用于制定施行令。

科学技术信息通信部由80多名民间专家组成了“下位法令整顿团”，并召开了70余次会议。2025年9月8日，在国家人工智能战略委员会第一次全体会议上，报告了施行令草案。同月17日，还公开了2件公告和5件指南的草案。施行令的立法预告于11月12日至12月22日进行，为期40天，期间听取了产业界、学术界和市民团体的意见。

2026年1月21日，施行令（总统令第36053号）予以公布。正文共至第33条，附则为2条。

高影响人工智能的判断标准设定为两个阶段的程序。第一阶段确认领域，包括能源、医疗保健、招聘、贷款审批、交通、公共服务、学生评估等领域。第二阶段评估是否“对人的生命、身体安全及基本权利产生重大影响或引发危险的担忧”。指南建议将判断依据以文件形式留存，若向科学技术信息通信部部长请求确认是否属于该范畴，将在基础的30日内得到答复；若有延长，则在追加的30日内得到答复。

高性能人工智能（大规模人工智能）的基准线也已确定。对象为学习累计运算量达到 10^{26} 次方FLOP以上，且应用了最前沿技术，并可能对生命、身体和基本权利产生广泛影响的系统。与EU AI Act以 10^{25} 次方FLOP为基准相比，这是高出一个数量级的门槛。韩国国内绝大多数经营者都不在这个区间内。

次日，即1月22日，法律和施行令同时开始施行。科学技术信息通信部在同一天发布了五大指南，分别是：确保透明性指南、确保安全性指南、高影响人工智能判断指南、经营者责任指南、人工智能影响评估指南。虽然没有法律约束力，但它们发挥着实务指示的作用，向经营者展示履行法律上义务的方法。2件公告（最前沿人工智能确保安全性公告、高影响人工智能经营者责任公告）也一并施行。

同一天，人工智能基本法支持中心开门运作。该中心由韩国软件产业协会（KOSA）运营，截至施行第10周的2026年3月31日，已受理了552件咨询。该中心提供匿名咨询服务，深入的法律审查最多需要14天。

4. 施行前已作修订：2026年1月20日修订法

韩国《人工智能基本法》原法将重心放在了安全与规制上。缺乏产业培育、公共领域引入、弱势群体支持等促进条款。第22届国会在原法施行的前两天公布了修订法。

2025年期间，崔敏姬、李正宪、张喆敏、崔宝允等议员代表提出了9项修订案。11月21日，信息通信广播法案审查小委员会进行了审查，经朝野协商制定了委员会替代法案；该法案经11月24日科广委全体会议，于12月30日通过了全体会议。2026年1月20日，作为第21311号法律予以公布。

修订内容共有8项。

国家人工智能委员会更名为国家人工智能战略委员会（第7条）。这并非仅是名称的变更。其政策协调和绩效管理权限得到了扩充。委员长由总统担任，委员定额在60人以内，3名副委员长中1名为全职。除了18名当然委员（16个中央行政机关首长、国家安保室第三次长、总统秘书室负责人工智能的首席秘书官）外，还有最多41名委任的民间委员；加上8个分科委员会的多达52名分科委员，总计形成了一个85人的体制。48个中央行政机关的人工智能最高责任官（CAIO）全员被指定为副部级，每月召开例行协议会。

新设了促进公共部门引入人工智能的条款（第16条第3款和第4款）。国家机关在购买产品或服务、发包劳务时，必须优先考虑人工智能。其中也加入了公务员免责条款。若无故意或重大过失，公务员无需承担因引入人工智能而产生的赔偿责任。从企业角度来看，这是在面向政府的市场（B2G）中开辟新需求的一种结构。

确立了设立人工智能研究所的法律依据（第22条之2）。大学或企业在获得科学技术信息通信部部长许可后即可设立。

确立了支持弱势群体的法律依据。修订法新引入了在利用人工智能方面遇到困难的弱势群体概念，并具体确立了对高费用人工智能服务获取提供费用支持的依据（第17条之2）。2026年7月20日公布的修订施行令将该范围界定为残疾人、65岁以上老年人、基础生活保障受领者及次上位阶层、求职者、农渔业人员、多文化家庭成员、脱北者、保有经历的女性、单亲家庭支持对象等。

此外，人工智能领域的创业支持、专门人才培养、公共数据作为训练数据提供的依据、人工智能技术教育支持也包含在了修订法之中。

施行日期分为两个阶段。战略委员会的改组及专门人才支持等条款自2026年1月22日起即时适用。而公共部门引入人工智能、人工智能研究所设立程序、弱势群体费用支持等需要整顿下位法令的条款，则于2026年7月21日施行。

5. 经营者的时间表：现在须遵守的与尚在缓期中的

有一些自2026年1月22日起即刻生效的义务。

即事先告知义务（第31条第1款）。在提供使用了高影响人工智能或生成式人工智能的产品或服务时，必须提前将该事实告知使用者。生成物标识义务（第31条第2款和第3款）也已开始实行。人工智能生成的成果物，必须按类型进行标识。对于深度伪造（Deepfake），仅使用不可见水印是不够的，必须加上肉眼可见的标识。明显能看出是人工智能生成物的情况，以及内部业务用途，则作为例外予以认可。

对于高影响人工智能经营者，五项责任集中在第34条第1款这一条款中被赋予。分别是：制定并运营风险管理方案（第1项）；针对最终结果及判断的主要标准以及训练数据概况，制定并实施说明方案（第2项）；制定并运营使用者保护方案（第3项）；确保人能进行管理和监督的体系（第4项）；编写履行依据文件并保管5年（第5项）。根据施行令，各方案的主要内容必须在主页上公布。

高性能人工智能经营者必须依据确保安全性义务（第32条），识别并评估风险，采取缓解措施，然后提交其结果。

若在韩国境内没有住所或营业所的海外人工智能经营者超过了一定标准（使用者人数、营业额等），则必须指定韩国境内代理人（第36条）。若不遵守，将面临3000万韩元以下的行政罚款。

行政处罚的处以（第43条）适用宽限辅导期。政府已表明方针，将至少缓期一年，直至2027年1月之后。对于遗漏透明性义务或文件化不完备等事项，优先采用行政指导和纠正劝告。但对于滥用于深度伪造犯罪、导致人命事故、大规模侵犯人权等重大事项，则可立即展开事实调查。

对于已经依据现有的《个人信息保护法》、《医疗器械法》、《信用信息法》等履行了同等措施的经营者，在相重叠的范围内，将被认可为已履行高影响人工智能经营者的责任（第34条第1款）。但此项认可并不适用于事先告知和生成物标识等透明性义务，也不适用于确保安全性义务。这是旨在减轻重复规制负担的机制。

6. 欧盟：起步早，却迟到

欧盟委员会正式发布人工智能法案提案书是在2021年4月21日。这是在韩国第21届国会提出首个法案的9个月之后。欧洲创下了全球率先提出全面性人工智能规制法案的纪录。在此前一年的2020年2月，欧盟委员会就已发布了《人工智能白皮书》，并确立了基于风险（risk-based）的方法路径。

在欧盟委员会提出提案后，历经了约3年零1个月的立法程序。2022年12月欧盟理事会采纳了共同立场；2023年6月欧洲议会采纳了谈判立场（赞成499，反对28，弃权93）。同月开启了三方谈判（trilogue），并在12月6日至9日长达3天的马拉松式谈判后达成了临时协议。随后，于2024年3月13日通过欧洲议会全体会议的最终表决（赞成523，反对46，弃权49），5月21日获欧盟理事会全票批准，7月12日登载于欧盟公报，最终于8月1日生效。

生效与施行并非一码事。EU AI Act采用的是分阶段施行结构。2025年2月2日，被禁止的人工智能系统（社会评分系统、潜意识操纵人工智能等）以及人工智能素养义务率先适用。欧盟法案中存在韩国所没有的内容，那就是源头禁止条款。像社会评分系统、无差别生物识别监控、潜意识操纵等类型，是完全不被允许的。2025年8月2日，通用人工智能（GPAI）模型相关的义务开始生效。由约1000名

利益相关者参与制定的GPAI行为准则，也在同一时期获得了适当性决定。透明性（文件化，保存10年）、著作权（尊重robots.txt）、安全与安保（系统性风险管理）是其三大支柱。

高风险人工智能义务原定于2026年8月2日适用。然而，在2025年11月19日，欧盟委员会提出了“数字综合（Digital Omnibus）”修订案。其核心在于延长期限。独立型高风险人工智能（附件III）的适用时间点，从2026年8月2日推迟到了2027年12月2日，延后了约16个月。内置于受规制产品中的高风险人工智能（附件I），其适用时间从2027年8月2日再推迟1年至2028年8月2日。人工智能规制沙盒运营义务也延期至2027年8月2日。修订案明示了将中小企业规制成本降低35%的目标，并增加了对未经同意生成亲密图像的人工智能以及生成儿童性剥削物的人工智能的新禁令。

该综合修订案于2026年5月7日达成临时协议，6月16日获欧洲议会正式采纳，6月29日经欧盟理事会最终批准。

韩国《人工智能基本法》自2026年1月22日起，已开始适用事先告知、生成物标识、高影响经营者责任等主要经营者义务。而EU AI Act的高风险人工智能义务直到2027年12月才会适用。率先提出法案的欧洲，在高风险人工智能经营者规制方面，反而落后韩国近两年，出现了逆转的局面。如此广泛且率先执行全面性人工智能法律中经营者义务的国家，韩国尚属首例。

两地在行政罚金上的差距极其悬殊。欧盟的3500万欧元（约550亿韩元）与韩国的3000万韩元之间，有着超过1000倍的鸿沟。欧盟采用的是强制要求进行事前适当性评估（conformity assessment）的事前控制模式；而韩国则采用了通过自主影响评估并与公共采购挂钩，来引导市场自愿遵守的模式。即使在高性能人工智能的运算量基准上，欧盟的 10^{25} 次方与韩国的 10^{26} 次方之间也存在十倍的差异，这使得规制对象的覆盖范围本身就截然不同。

7. 中国：先出规定，综合法在后

中国并没有一次性制定全面的人工智能法律。其采取的模式是，先出台按领域的具体规定并予以执行，然后在后端筹备综合性法律。

2022年3月1日，《互联网信息服务算法推荐管理规定》开始施行。该规定由国家互联网信息办公室（CAC）、工业和信息化部、公安部、国家市场监督管理总局四个部门联合制定，内容涉及算法推荐服务的透明性以及使用者的选择权。2023年1月10日，《深度合成管理规定》开始施行，强制要求对合成内容加上肉眼可见的标识。同年8月15日施行的《生成式人工智能服务管理暂行办法》，规定了遵守“社会主义核心价值观”的义务，以及对具有舆论属性服务的安全评估机制。煽动颠覆国家政权、恐怖主义、民族仇恨等内容均被禁止。而韩国和欧盟的法律中并未包含此类内容控制条款。

事前备案制度是核心的执法手段。具有舆论属性或社会动员能力的服务，必须在发布后10个工作日内进行算法备案。截至2025年11月，已有611个生成式人工智能服务完成了备案，算法备案总数超过了5000件。

2025年9月1日，《人工智能生成合成内容标识办法》开始施行，赋予了双重标识义务（显式标签与元数据或隐式水印）。甚至制定了强制性的国家技术标准（GB 45438-2025）。2026年7月15日，《拟人化人工智能交互服务暂行办法》将开始施行。此前的5月份，还以政策文件的形式发布了关于人工智能智能体规范应用的实施意见。其中限制向未成年人提供虚拟亲密关系服务，并要求达到一定规模以上的服务必须进行安全评估。

执法行动已经相当猛烈。仅在2024年上半年，重庆就实施了11件行政处罚，关闭了142个网站，对101个平台进行了约谈整改。2025年2月，国家互联网信息办公室下架了未经安全评估的生成式人工智能服务应用和小程序。在2025年6月上海的“黄浦亮剑”行动中，删除了82万件违法内容，并使约2700个人工智能智能体停用。

全面性的人工智能法尚未出台。尽管在2017年国务院的《新一代人工智能发展规划》中宣告要在2025年前完善人工智能法律法规体系，但截至2026年7月，综合法的草案尚未正式公开。2024年3月，中国法学界的研究团队公布了一份学者建议稿（共9章96条），乔治敦大学安全与新兴技术中心（CSET）公开了其英文翻译版。2026年5月，国务院在年度立法计划中表示，将“加快推进人工智能健康发展的综合性立法”。预计内容将包括记录所有训练数据来源、对超过千亿参数的模型进行强制安全影响评估、在做出影响合法权利的决定时要求人类介入，以及高达5000万元人民币（约95亿韩元）的罚款等。预计全国人大将于2027年底批准，并于2028年施行。

三个国家的规制强度排序鲜明，依次为欧盟、中国、韩国。仅看最高制裁金额，以现行法为基准，欧盟为3500万欧元，韩国为3000万韩元；而中国有预测称，一旦综合法制定出台，最高罚款将达5000万元人民币（约95亿韩元）。在规制方式上也各有不同。欧盟采用全面规制法，中国采用分领域个别规定与内容控制相结合的方式，而韩国则是振兴法与规制法相混合。中国这种“法律加标准”的双轮驱动模式，即由国家标准化管理委员会（TC260）制定软性规范，而后转化为硬性规范的体系，在韩国和欧盟是找不到对应机制的。

8. 552个疑问

2026年7月21日，修订法的剩余条款开始施行。公共部门引入人工智能优先考虑、人工智能研究所设立程序、弱势群体费用支持等均具备了法律效力。修订后的施行令历经了2026年5月21日至6月19日的立法预告，于7月20日作为总统令第36506号予以公布，并于7月21日起施行。

在法律施行的当天，即2026年1月22日，国会立法调查处的郑俊和（音译）研究官发表了《实现国家人工智能战略的行政课题》。文章指出，高影响人工智能概念与要件的具体化，以及按经营者类型区分义务，是留下的未解难题。韩国法制研究院的郑元俊（音译）研究委员在其2025年9月的报告中也指出，需要遵守法律的当事人（守法主体）的范围尚不明确。

人工智能基本法支持中心在法律施行仅10周后，便涌入了552件咨询。法律虽然已经通过，施行令业已出台，指南也分发完毕，但经营者们依然在提问：我们公司的人工智能服务是否属于高影响范畴？事先告知究竟要以何种形式进行？海外合作伙伴指定韩国境内代理人的标准是什么？在“有法可依”与“能够守法”之间，堆积着整整552个疑问。

第二章. 我是该法的适用对象吗：人工智能经营者的定义与适用范围

2026年1月22日，是《人工智能基本法》施行的日子。首尔某大学医院的影像医学科。专科医生将胸部CT影像显示在显示器上。屏幕一侧的AI辅助阅片系统用红圈标出疑似肺结节的部位。这位医生每天与AI一起阅片数十次。这家医院是《人工智能基本法》第2条第7项所说的人工智能经营者吗？在韩国法律中，不是。因为它只是接受AI阅片系统并用于诊疗，因此这家医院是第8项的“使用者”。它不承担法律上的义务。如果同一位医生在慕尼黑的医院使用同一个系统，那么在欧盟《人工智能法案》中，该医院就是部署者（Deployer）。它必须按照使用指南运行系统，安排人工监督，并保存系统运行日志。同样的技术和同样的行为，法律地位却截然不同。

1. 法律视角下的人工智能

《人工智能基本法》的正式名称是《关于促进人工智能发展与构建信任基础等的有关基本法》。法律第20676号。该法于2024年12月26日获得国会全体会议通过，于2025年1月21日制定并公布，并于2026年1月22日施行。在施行前两天的2026年1月20日，第2条等条款被修改，而施行令（总统令第36053号）于2026年1月21日公布，2026年7月20日被修改，并于次日施行。

第2条包含了十二个定义。对经营者而言最核心的有六个。

第1项定义了“人工智能”。即“通过电子方式实现人类所具备的学习、推理、感知、判断、语言理解等智力能力的事物”。该定义列举了人类的五种认知功能，并囊括了以电子方式模仿这些功能的全部技术。学术界评价，该定义是一种将人工智能比拟为人类能力来加以描述的拟人化进路。

欧盟《人工智能法案》第3条第1款的进路有所不同。其以“基于机器的系统（machine-based system）”一词作为出发点，描述了一种接收输入进行推理，

并输出预测、推荐或决定等结果的系统。韩国是以人类能力为标准来看待；欧盟则是以机器功能为标准来看待。

第2项单独定义了“人工智能系统”。即“具有不同程度的自主性与适应性，为实现特定目标，推理出预测、推荐、决定等会对现实及虚拟环境产生影响之结果的基于人工智能的系统”。如果说第1项的人工智能是技术本身，那么第2项的人工智能系统就是包含该技术并正在运行的体系。它必须具有自主性，具有适应性，并能够通过推理输出结果。

韩国法律采取了这种将“人工智能”与“人工智能系统”分离开来的二段式结构。欧盟《人工智能法案》则统一将其定义为“AI系统”。

这种区分在实务中能够立即派上用场。物流公司在仓库中使用的RPA（即基于规则的重复性自动化程序）并非人工智能系统。按照预先设定的规则重复执行的软件既无自主性，也无适应性和推理能力。相反，如果该物流公司使用优化配送路线的机器学习模型，那么它就是人工智能系统。它会学习数据，适应环境变化，并通过推理推荐路线。

2. 开发经营者、使用经营者、使用者

该法第2条第7项将人工智能经营者分为两类。

a目的开发经营者是指“开发并提供人工智能者”。b目的使用经营者是指“利用a目之经营者所提供的人工智能，提供人工智能产品或人工智能服务者”。第8项的使用者是指“接受人工智能产品或人工智能服务者”。

其核心在于“提供”这一动词。开发经营者提供人工智能本身。使用经营者利用该人工智能来提供产品或服务。使用者则只是接受提供。

OpenAI开发GPT模型并以API形式提供。它是开发经营者。韩国国内一家金融科技公司调用该API制作了客户咨询聊天机器人，并将其交付给银行。它是使用经营者。银行接受该聊天机器人并将其用于接待客户。如果仅用于内部业务，它

更接近于使用者。然而，即使是成品聊天机器人，如果银行以自己的名义提供给客户，则需要考量其是否属于第2条第7项b目的使用经营者。

这就是韩国法律的设计选择。它没有对引入AI并将其用于本职业务的机构施加义务。

不问其是营利性质还是非营利性质。无论是法人、团体、个人还是国家机关，只要是从事与人工智能产业相关的经济活动者，就会成为适用对象。如果大学实验室制作了AI模型且仅用于论文研究，则不是经营者。如果将该模型作为API开放并持续向外部提供，则是经营者。

可以同时具备两种地位。如果一家金融公司开发了自有的AI信用评估模型并提供给关联银行，同时又引进外部AI用于内部贷款审批，那么对于信用评估模型而言，它是开发经营者，而对于贷款审批AI而言，它则是使用经营者。

开发经营者承担第一责任。这是对模型的设计、训练及安全性承担的源头责任。使用经营者承担第二责任。它必须在自身服务的语境中告知透明度情况，并在应用于高影响领域时制定风险管理方案。如果开发经营者已经履行了关于风险管理、说明和使用者保护的措施，且使用经营者没有大幅改变人工智能系统的原本目的或用途，那么就这三种措施而言，可视为使用经营者也已履行。这是防止双重负担的机制。

微调模糊了这一边界。

有些公司会引入外部模型，用自身数据进行微调，然后将结果作为新服务提供。它们并没有从头开始制作原本的模型。但它们也不是原封不动地使用。在法律条文“开发并提供人工智能者”的表述中，微调后提供服务的情形有符合该描述的余地。如果微调是实质性改变模型行为的行为，并且公司以自己的名义推出该结果，那么这家公司很有可能被归类为开发经营者。截至2026年7月，科学技术信息通信部尚未出台官方权威解释。不确定性依然存在。

《人工智能基本法》援助中心在法律施行后的十天内接到了172件咨询。其中，透明度义务占56.4%，确认高影响人工智能占17%，定义问题占10.6%。这意味着经营者们在该法面前都在问“我是适用对象吗”。化妆品公司用AI提供广告服务时到底是经营者还是使用者，信用评估公司向金融机构提供AI模型时到底哪部分是规制对象，在实务中积累了许多界线不分明的情況。

3. 生成式、高影响、高性能

在人工智能经营者这一框架之上，还叠加了三种分类。

第2条第5项，生成式人工智能。即“模仿输入数据的结构与特征，生成文字、声音、图像、视频及其他各种结果物的人工智能系统”。ChatGPT、Midjourney、Suno AI均属于此类。提供生成式人工智能的经营者会增加两项义务。即必须在生成物上标注由AI制作的生成物标识，并须向使用者事先告知。

第2条第4项，高影响人工智能。即“恐将对人的生命、身体的安全性及基本权利产生重大影响或引发风险的人工智能系统”，且它被用于法律所列举的领域。

共有十一个领域。能源供应，饮用水生产流程，保健医疗体系，医疗器械及数字医疗器械，核物质与核设施的安全管理，这是前五个。为犯罪侦查或逮捕职务而进行的对人脸、指纹、虹膜等生物识别信息的分析与应用，像招聘或贷款审批一样对个人权利与义务产生重大影响的判断，交通工具、交通设施及交通体系的主要运作，国家、地方自治团体及公共机构的资格审查与费用征收，从幼儿教育到中等教育的学生评价，这是另外五个。最后一个是可以通過总统令追加的开放式项目。

判断是否为高影响的结构分为两步。是否被用于相关领域；以及，是否对生命、身体安全性或基本权利产生重大影响。韩国法制研究院指出，在这一判断结构中，“影响”和“风险”在法律中被并行使用，且“安全性”和“基本权利”由“及（and）”连接，从而造成了解释上的混乱。

高性能人工智能并未出现在基本法第2条中。它根据基本法第32条的授权，在施行令第24条中予以规定。标准有三项，并且必须全部满足。用于训练的累计运算量须达到 10^{26} 次方FLOP及以上；须应用目前正在被使用的人工智能技术中最前沿人工智能技术进行构建和运行；其风险程度恐将对人的生命、身体的安全性及基本权利产生广泛且重大的影响。并不是只要迈过运算量这一个门槛就能成为适用对象。符合该标准的人工智能系统的经营者，将被赋予在整个生命周期内识别、评估和缓解风险，构建监测和应对安全事故的风险管理体系，以及提交其履行结果的义务。

其运算量门槛比欧盟高出十倍。欧盟《人工智能法案》作为具有系统性风险的通用AI模型标准的运算量是 10^{25} 次方FLOP。韩国则定为了 10^{26} 次方。这可以看出其试图将规制对象狭窄限定的意图。

4. 跨越国界的法律

该法不仅适用于韩国境内。

第4条第1款规定：“即使是在国外实施的行为，若对韩国国内市场或使用者产生影响，亦适用该法。”若提供韩语服务，或以韩国消费者为对象进行营业，或以任何形式对韩国国内市场产生影响，该法便会随之适用。

将总部设在海外的AI经营者中，达到一定规模以上的，须指定韩国境内代理人。根据第36条的授权，施行令第29条第1款规定的标准有四项，只要满足其中一项就会产生义务。上一年度销售额达到1万亿韩元及以上。人工智能服务部门上一年度销售额达到100亿韩元及以上。最近3个月韩国国内日均使用者人数达到100万人及以上。以及因未履行中止命令或纠正命令而曾被处以行政罚款的经营者。韩国境内代理人代替经营者执行提交确保安全性义务履行结果、请求确认是否属于高影响人工智能的工作，并负责提供履行高影响责任所需的援助。

欧盟《人工智能法案》第2条将域外适用细分为七种类型。即向欧盟境内投放AI的提供者（provider），位于欧盟的部署者（deployer），其输出内容在欧盟境

内被使用的第三国经营者，进口商，分销商，将AI整合进自身产品的制造商，以及欧盟境内受影响的人（affected person）。中国没有明示域外适用的条款。仅写明适用于“向中国境内公众提供服务的情形”，且以技术屏蔽代替了明示的域外适用，以此来确保管辖权。

韩国用一句话进行了概括。欧盟则按类型进行了拆分。中国更是没有明示条款本身。

5. 军事目的之例外

第4条第2款规定了例外情况。

“该法对于仅出于国防或国家安全目的而开发、利用的人工智能，且属于总统令所规定之人工智能者，不予适用。”

“仅出于……目的”一词是核心。英文为exclusively。只有专为国防打造且专为国防使用的人工智能才属于此例外。同时针对军事和民用的军民两用（dual-use）人工智能不属于例外。它仍适用该法。

施行令第2条限定了具体的范围。国防部长官指定的国防信息通信网建设与武器体系运行。国家情报院长指定的预警系统、安全课题、反恐活动、核心技术泄露调查。警察厅长指定的反恐与尖端战略技术侵害侦查。该范围之外的人工智能，即使是军队或情报机构使用，也适用该法。

欧盟《人工智能法案》第2条第3款也使用了相同的结构。“exclusively for military, defence or national security purposes.”两者都使用了排他性的限定词，其结论都是对军民两用情形予以规制，这一点是相同的。如果说有差异，那就是欧盟的条款无需单独的下位法令即可直接适用，而韩国则将具体范围授权给了总统令。

6. 名为欧盟的镜子

韩国《人工智能基本法》和欧盟《人工智能法案》对同一个问题给出了不同的答案。

从经营者分类开始就不同。欧盟区分提供者（Provider）和部署者（Deployer）。韩国则区分开发经营者和使用经营者。名称看似相似，但范围却大相径庭。

欧盟的提供者（Provider）是指开发AI系统或通用AI模型，或者委托开发后以自己的名义或商标将其投放市场者。其明确指出了“以自己的名义或商标（under its own name or trademark）”，因此连白标（white-label）方式的委托AI也囊括在内。在韩国的开发经营者定义中，并没有与此相对应的规定。

部署者（Deployer）和使用经营者的差异更大。欧盟的部署者（Deployer）是指在专业语境下，在自身权限范围内使用AI系统的所有自然人或法人。仅排除了个人的非专业使用。范围很广。韩国的使用经营者则是指利用开发经营者提供的人工智能来提供产品或服务者。它以供应关系为前提，并且必须对外提供。范围很窄。

我们再来看看首尔某大学医院的影像医学科。拍完肺部CT后，AI首先会阅读影像并在疑似结节处留下标记。医生参考该标记撰写阅片报告，并且以医生的名义出具诊断。同样的设备、同样的软件也安装在德国慕尼黑的医院里。机器虽然一样，但两家医院的法律地位却不同。

在欧盟，该医院是部署者（Deployer）。根据欧盟《人工智能法案》第26条，其必须遵守制造商的使用指南，必须指定具备监督权限与能力的人员来监视AI，并且必须将系统自动生成的运行记录保存至少6个月。在韩国，同样的医院是使用者。《人工智能基本法》没有对使用者施加任何义务。

为什么会产生这种差异呢。答案在定义条款中。人工智能使用经营者是指利用开发经营者提供的AI来提供“人工智能产品或人工智能服务”者。这并非适用于任何服务，而是向他人推出的东西本身必须是AI产品或AI服务。医院向患者推出

的是诊疗。患者接受的并不是AI服务，而是参考了AI的医生诊断。AI只是在医院内部用完即止的工具。因此，医院不是提供AI者，而是接受提供者，即使用者。这与引进贷款审批AI并将其用于审批的银行停留在使用者的地位是同样的道理。

不过，这里有一条界线。判断标准在于交到患者手中的东西是什么。如果是医生的诊断，医院就是使用者；如果是AI的输出结果，医院就是使用经营者。如果医院在患者用应用程序中附加了AI问诊聊天机器人，让患者直接与AI对话，或者未经医生审核便将AI阅片结果通知给患者，那么患者接受的就是人工智能服务本身。此时，医院就该服务而言便成为了使用经营者，随之而来的还有确认是否属于高影响人工智能、采取旨在确保安全性与可信度的措施以及事先告知义务。我们必须记住，地位并不是附属于机构层面，而是附属于服务层面。原本在影像医学科是使用者的医院，在开放患者用聊天机器人那一刻起，就该聊天机器人而言便成为了使用经营者。

这种结构是立法上的选择。欧盟从医院将AI作为工具使用的阶段便开始施加义务，而韩国则是到了医院向患者推出AI的阶段才开始施加义务。规制介入的门槛位置不同。韩国法律减轻了引入AI之机构的负担。科学技术信息通信部亲自发布指南时亦表示，这是将确保人工智能可信度的立法宗旨与减轻企业负担结合起来考量的结果。

在实务中请这样判断：若要医院维持使用者的地位，请在内部规定中明示由医生做出最终判断的程序，并将其记录在诊疗记录中。如果计划开放直接面对患者的AI服务，请不要在开放服务的当天，而是从准备之日起便具备使用经营者的义务。门槛虽然低，但在跨越的那一瞬间，所背负的担子绝不比欧盟的部署者（Deployer）轻。

欧盟将规范对象分为六种类型。即提供者（Provider）、部署者（deployer）、进口商（importer）、分销商（distributor）、授权代表（authorized representative）、产品制造商（product manufacturer）。将这六者统称为运

营者（operator）。韩国则只有两种。即开发经营者和使用经营者。欧盟在供应链的每一个环节均赋予了量身定制的义务，而韩国则将开发和使用这两大块作为运营方式。韩国法制研究院指出，这种结构过于宽泛，以至于连政府机构也被纳入到人工智能经营者之中。

风险分类体系也不同。韩国使用“影响（impact）”一词。这是高影响人工智能这一单一范畴。欧盟则使用“风险（risk）”一词，并将其分为四个等级。即被禁止的惯例（unacceptable risk）、高风险（high-risk）、有限的风险（limited risk）、极小风险（minimal risk）。韩国没有禁止范畴。欧盟从源头上禁止了诸如社会评分等惯例。以执法为目的的实时远程生物识别原则上被禁止，仅在搜寻失踪人员、应对紧迫威胁、锁定重大犯罪嫌疑人等狭窄的例外情况下，并在严格的审批程序下予以允许。韩国看重的是影响。欧盟看重的是风险。

欧盟对高风险AI施加了强制性的合格评定（conformity assessment），而韩国则将影响评估设为尽力义务。虽然风险管理等高影响责任在法律上属于履行义务，但即使违反也不会立即附带罚则，而是要到了未履行纠正命令的阶段，才会处以行政罚款。强制程度不同。欧盟举起了刀，韩国则递出了指南。

欧盟有通用AI模型（General-Purpose AI Model, GPAI）的概念。韩国则没有。韩国虽然定义了“生成式人工智能”，但它仅限于生成内容的系统。分类、推理、工具使用等不生成内容的通用模型是否属于该定义，目前尚不明确。

7. 中国，针对不同技术的不同法律

中国走了一条截然不同的道路。

它没有制定一部涵盖全部人工智能的单一法律。每当出现新的技术时，它便出台符合该技术的规定。2022年3月的《互联网信息服务算法推荐管理规定》，2023年1月的《互联网信息服务深度合成管理规定》，2023年8月的《生成式人工智能服务管理暂行办法》。在两年内出台了三部规定。

《互联网信息服务算法推荐管理规定》将算法分为五种类型。生成合成类、个性化推送类、排序精选类、检索过滤类、调度决定类。对于具有舆论属性或者社会动员能力的服务提供者，赋予其算法备案义务。须在提供服务之日起十个工作日内，将算法类型、应用领域、自评估报告等信息向网信部门备案。

《互联网信息服务深度合成管理规定》将深度伪造技术列举了六个子类型。文本生成与编辑、语音属性转换、非人脸图像与视频生成、人脸生成与替换、三维重建、沉浸式内容生成。该规定中出现了一个韩国或欧盟都没有的概念。即技术支持者。将不直接提供服务但提供技术支持者作为单独的适用对象。在韩国法律中没有与此概念对应的定义，在欧盟《人工智能法案》中也没有。

《生成式人工智能服务管理暂行办法》将“生成式人工智能技术”定义为“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”。这与韩国的生成式人工智能定义相似，但韩国一直描述到“模仿输入数据的结构与特征”这一机制，而中国仅凭输出能力进行定义。

中国的经营者分类为提供者和使用者。提供者是指利用生成式AI技术提供服务的组织或个人，也包括API的提供。使用者则是指使用该服务生成内容的组织或个人。韩国的使用经营者与中国的使用者在名称上相似，但韩国的使用经营者是指对外提供产品或服务者，而中国的使用者则是指生成内容者。范围和义务也均不相同。

中国没有涵盖整个人工智能的统一的定义。各个规定只定义属于自己管辖的技术。也没有域外适用条款。只写明适用于“向中国境内公众提供服务的情形”。以技术屏蔽代替了明示的域外适用，以此来确保管辖权。

在中国规制中引人注目的是备案制。韩国倾向于事后自律规制。欧盟是在事前合格评定中加上了被禁止的惯例。中国则是将备案与内容控制结合在一起。这是一种由规制机构通过备案进行掌握，并在事后进行控制的结构。

8. 三个国家的制裁

法律的牙齿在于制裁。

欧盟《人工智能法案》的罚款上限，在违反被禁止的惯例时，为3,500万欧元或上一财年全球销售额的7%中的较高者。由于全球大型科技企业适用按销售额比例计算，因此实际罚款可达数千亿韩元。

韩国《人工智能基本法》中没有罚款。对经营者违反义务的金钱制裁仅有第43条的行政罚款，其上限为3,000万韩元。约合2万美元。在欧盟的3,500万欧元与韩国的3,000万韩元之间，存在着四位数的差距。

中国并没有设置一个贯穿三部规定的单一罚款上限。《互联网信息服务算法推荐管理规定》本身设定的罚款在1万元至10万元人民币之间，而《生成式人工智能服务管理暂行办法》则是将警告和纠正命令放在首位，金钱制裁则交由《网络安全法》等个别法律处理。但是，中国有着停业整顿和暂停服务这样的制裁。对于试图在中国市场继续开展业务的企业而言，暂停服务命令比任何金钱制裁都要沉重。

制裁的差距反映了规制的哲学。

欧盟将保护基本权利放在首位。如果风险很高，就要在事前阻止，这是一种预防性进路。设定禁止惯例，对高风险AI施加强制性合格评定，并对最严重的违规行为处以最高可达销售额7%的罚款。

韩国则将促进创新与保护权利并列。该法第1条将“促进人工智能发展”与“构建信任基础”并列写入，便是明证。没有禁止范畴，合格评定是自律性质的，且行政罚款上限为3,000万韩元。学术界将该法评价为骨架法。国会立法调查处则指出其对消费者保护不足。该法的设计是侧重于创新的，那是故意的选择。不过，在“3,000万韩元的行政罚款能否改变全球AI企业行为”的疑问面前，这一选择的代价尚未被计算出来。

中国将社会稳定与内容控制放在首位。垂直划分的规制结构是指每当出现新的技术时，便迅速出台符合该技术的控制规范的方式。在三个国家中，论规制速度，中国走在了前面。

话题再次回到首尔那家大学医院的影像医学科专科医生。这位医生在韩国是使用者，在慕尼黑是部署者（Deployer）。在北京则没有合适的称呼。因为CT阅片辅助系统不生成内容，所以在针对不同技术分别制定的中国的三部规定中，没有任何一处是完全贴合的位置。各个国家对同样的行为冠以不同的名称，或者干脆不设名称，并施加不同重量的义务。在《人工智能基本法》给予产业喘息空间的同时，欧盟筑起了基本权利的围栏，而中国则撒下了备案的大网。您是该法的适用对象吗？翻开该法第2条第7项便可看到答案。如果开发并提供人工智能，或者利用其提供产品或服务，那么就是经营者。如果仅仅是接受并使用，那么就是使用者。而且，该地位不是附属于机构，而是附属于服务层面。原本在影像医学科是使用者的医院，在开放患者用聊天机器人那一刻起，就该聊天机器人而言便成为了使用经营者，而在进行微调的瞬间便有可能越界走向开发经营者那一边。这披着基本法外衣的法律，在下位法令、权威解释与案例的积累中究竟会具备怎样的厚度，目前仍未有定论。

第二部 自我诊断：我们的服务适用哪些规制

第3章. 第一阶段判定：我们是生成式人工智能经营者吗

2026年2月，首尔江南的一家电子商务公司。在营销团队的会议室里，团队负责人指向屏幕。AI在夜间自动撰写了上传至自家购物网站的3000件新商品的商品描述。只要放入照片，就会出现描述材质、尺寸和穿着效果的句子。法务团队发来联络。根据《人工智能基本法》第31条第2款的规定，生成式人工智能生成的成果物应当进行标识。这3000件商品描述是否全都要加上“本文由AI撰写”？团队负责人首先需要确认的是，自家的服务是否属于法律所称的“生成式人工智能”。

1. 法律条文的三大条件

第2条第5项：“通过模仿输入数据的结构和特征，生成文本、声音、图像、视频及其他各种成果物的人工智能系统。”

将这一定义进行拆解，可以得出三个条件。

必须“模仿”输入数据的结构和特征。这是指模仿AI模型所学习数据的模式和结构，从而创造出新成果物的机制。典型代表是大语言模型在学习了数十亿条文本后，模仿其句子结构和词汇模式来创造新句子。

成果物的类型必须是“文本、声音、图像、视频及其他各种成果物”。法律明确列举了文本、音频、图像、视频，并通过“及其他”保持了开放性。代码、3D模型、分子结构等形式也可以包含在内。

必须是第2条第2项所定义的人工智能系统。它应当是一个具备自主性和适应性，并能通过推理得出成果物的系统。仅按照预设规则进行重复操作的软件不属于人工智能系统。

only这三个条件全部满足，才属于生成式人工智能。缺一不可。

2. “模仿”这一关键

在这三个条件中，解释空间最广的是“模仿输入数据的结构和特征”。

大语言模型（LLM）通过学习数十亿条文本来掌握标记（token）的概率分布。在生成新句子时，它会结合前面的标记来计算下一个标记的概率，并逐个输出。它虽然模仿学习数据的句子结构和词汇模式，但并不会原封不动地重现学习数据。这就是模仿。

生成对抗网络（GAN）学习真实图像的统计分布，并从中提取新图像。扩散模型（Diffusion Model）通过学习对图像添加噪声然后再去除噪声的过程来制作新图像。这两种方式都是捕捉学习数据的结构性特征，并生成具有相似属性的成果物。

模仿不是复制。原封不动地重复学习数据只属于搜索或数据库查询。在学习了数据所具有的模式、分布和规律性后，以此为基础创造出此前不存在的事物的过程才是模仿。

这一要件具有过滤掉基于规则的系统的功能。例如一个仅在预设模板中插入变量来生成句子的系统：“○○您的订单号XXXX已受理。”这并不是模仿数据的结构，而是按照程序员编写的规则运行。它不属于生成式人工智能。

3. 显而易见的案例

ChatGPT是生成式人工智能。它学习了大量文本，模仿文本的结构和特征，生成新的文本。

Midjourney也是如此。它学习了大量图像，模仿图像的结构和特征，生成新的画作。

Suno AI是声音方面的代表。它学习了大量音源数据，模仿音乐的结构和特征，生成新的音乐。

OpenAI的Sora是视频方面的代表。它学习了大量视频数据，模仿视频的结构和特征，生成新的动态视频。

GitHub Copilot用于编写代码。它学习了大量源代码，模仿代码的结构和特征，生成新的编程代码。代码属于法律所称的“其他各种成果物”。

运营这些服务的经营者是生成式人工智能经营者。需要承担生成物标识和使用者事先告知的义务。

4. 边界之上的案例

困难的并不是那些显而易见的案例。

我们先来看推荐算法。Netflix分析观看历史，向您呈现“您可能喜欢的电影”。酷澎（Coupang）基于购买记录，向您展示“推荐商品”。推荐算法是从已有的内容中挑选并进行展示。电影列表早已存在，算法只是决定将哪部电影展示给谁，并不创造新的内容。因为这属于“筛选已有项目”而非“生成成果物”，所以不属于生成式人工智能。

但如果推荐算法同时撰写了个性化的文案，情况就不同了。如果AI生成了诸如“如果您喜欢上周观看的悬疑片，我们也为您推荐这部作品”之类的定制化文案，那么这一文案生成功能就可能被纳入生成式人工智能的范围。这是筛选和生成融合在同一个服务中的情况。推荐功能本身虽然不属于生成，但制作推荐文案的功能可能属于生成。

再来看搜索自动补全。如果在NAVER或谷歌搜索框中输入“人工智能基”，系统就会推荐“人工智能基本法”。大多数自动补全功能只是显示过去输入过的搜索词，而不是创造新的文本。因此，很难将其视为生成式人工智能。

而搜索引擎结合AI以摘要形式展示搜索结果的情况则不同。谷歌的AI Overview和NAVER的AI搜索简报综合了多个搜索结果，重新生成了此前不存在的文本。这一功能属于生成式人工智能。在同一个搜索服务中，自动补全被划分为非生成，

而AI摘要被划分为生成。

我们来看翻译引擎。DeepL将英文合同翻译成韩文，Papago将日文论文翻译成韩文。翻译结果是文本。基于神经网络的机器翻译（NMT）学习了大量的双语平行语料库，通过模仿语言的结构特征来生成另一种语言的文本。这一过程可以契合“通过模仿输入数据的结构和特征来生成文本”的行为。不过，对于将翻译视为“转换（transformation）”还是“生成（generation）”，目前尚未形成确切的解释。翻译引擎需要有原文这一输入，并将其含义用另一种语言进行重构以输出新文本。从没有原文就没有成果物的角度来看，它与纯粹的生成性质不同，但从神经网络模仿语言模式来组装句子的角度来看，两者的机制是一致的。截至2026年7月，科学技术信息通信部尚未对这一边界给出官方解释。

我们来看AI美颜修图滤镜。当智能手机相机应用拍摄照片时，AI会自动调整亮度、色彩和清晰度。如果只是在保持照片基本结构的同时调整像素值，很难认为“生成”了新的图像。但如果AI将背景完全更换或改变人物的外貌，那就更接近于创造出此前不存在的图像。修图的程度越深，就越会跨越生成的边界。

边界不在于技术的名称，而在于成果物的性质。如果只是挑选已有的内容，则不属于生成。如果是对已有内容进行微调，是否属于生成则尚不明确。如果是创造出此前不存在的新内容，那就是生成。

5. 向生成式经营者追加的义务

提供生成式人工智能的经营者需要额外承担两项义务。

第一项是生成物标识义务（第31条第2款）。必须在AI生成的成果物上标明该事实。对于与实际情况难以区分的虚拟声音、图像、视频（即所谓的深度伪造类成果物），要求以使用者能够清晰识别的方式进行告知或标识（第31条第3款）。

《确保透明度指南》引导将此类成果物上带有肉眼可见的显性标识作为原则，因此仅凭隐性水印很难满足这一要求。标识方法的具体细节由《指南》分类型进行引导，而《施行令》对于显而易见属于AI生成物的情况（例如AI服务平台内的成

果物）以及以内部业务为目的的生成物承认例外。

第二项是使用者事先告知义务（第31条第1款）。拟提供利用高影响人工智能或生成式人工智能的产品或服务时，必须预先向使用者告知该产品或服务是基于该人工智能运行的事实。事先告知并非仅是生成式人工智能的义务。它也适用于利用高影响人工智能的产品和服务。但在生成式人工智能的情况下，事先告知与生成物标识相结合，使透明度负担又加重了一层。

这两项义务均自2026年1月22日施行之日起开始适用。虽然在处以行政罚款方面给出了至少一年的过渡期，但在滥用深度伪造进行犯罪或造成人身伤亡等重大案件中，可以立即展开事实调查。

生成物标识本身并不是目的。其目的是为了让使用者能够知晓现在所看到的文本、所听到的声音、所看到的图像是由人制作的还是由机器制作的。在施行初期的10天里，《人工智能基本法》支持服务台共受理了172件咨询，其中与透明度相关的咨询占比最大，这意味着该义务的范围和方法对经营者而言仍显陌生。

6. 欧盟关注的焦点：通用性

韩国《人工智能基本法》中包含了“生成式人工智能”。欧盟《人工智能法案》中则有“通用人工智能（General-Purpose AI，简称GPAI）”。这两个概念看起来相似，但分类标准不同。

韩国看重的是输出物。是否“生成”文本、声音、图像、视频是判断的标准。焦点在于创造出了什么。

欧盟看重的是通用性。欧盟《人工智能法案》第3条第63项将“表现出显著通用性（significant generality）并能执行广泛的各种不同任务”的AI模型定义为GPAI。其标准不在于创造出了什么，而在于能够执行多么广泛的任务。

这一差异造成了空白。

假设存在一个执行分类工作的通用AI模型。它判别电子邮件是否为垃圾邮件，对文档的主题进行分类，并识别图像中的对象。该模型输出的是“分类”结果，而不是重新创造文本、声音或图像。按照欧盟的标准，由于其能执行多项任务，因此可能属于GPAI。而按照韩国的标准，由于其不“生成”成果物，因而很难属于生成式人工智能。

执行推理和工具调用的AI智能体也会遇到同样的问题。管理使用者日程、分类电子邮件、查询数据库的智能体虽然有输出文本的部分，但其核心功能是调用外部工具并做出判断。目前尚不明确该智能体是否能完全被韩国法律中的生成式人工智能所吸纳。但在欧盟，可以将其作为构建在通用模型之上的系统（GPAI system）予以吸纳。

韩国法律中没有与GPAI相对应的概念。虽然存在“生成式人工智能”和“高性能人工智能”（这并非直接出现在法律中的名称，而是为了方便起见，对同时满足《施行令》第24条规定的三个要件——即累计学习计算量达到10的26次方FLOPs以上、应用最前沿人工智能技术、以及担心会对生命和人身安全及基本权利产生广泛而重大影响的人工智能系统的统称），但法律中并不存在以“通用性”为标准的分类。欧盟对GPAI模型和GPAI系统进行了区分，并建立了一套细致的体系，对存在系统性风险（systemic risk）的GPAI施加额外义务。而韩国则仅凭输出类型这一单一标准来分类生成式人工智能，在此基础上再以计算量和风险度为标准来划分高性能人工智能。两者的衡量标准数量是不同的。

义务的范围也存在很大差异。欧盟对GPAI提供者施加的义务已自2025年8月2日起开始适用。其中包括技术文档化和关于学习内容的公开摘要。约有1000人参与制定的《GPAI行为守则》是作为遵守该义务的手段而制定的。行为守则的三大支柱是透明度（文档化和保存10年）、版权（尊重robots.txt和最小化侵权）、安全与保障（系统性风险管理）。而在韩国，对生成式人工智能经营者施加的义务只有生成物标识和事先告知这两项。

韩国是对输出成果物的服务施加标识义务。欧盟则是以模型的通用性和系统性风险为标准来层层设定义务。监管的焦点不同。韩国侧重于告诉使用者“这是由AI制作的”。欧盟则在通用人工智能方面，侧重于管理模型本身的安全性和透明度。

7. 中国：以能力定义，用备案进行管控

中国《生成式人工智能服务管理暂行办法》（2023年8月15日起施行）将“生成式人工智能技术”定义为“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”。

这与韩国的定义相似，但也存在差异。韩国在定义中加入了“模仿输入数据的结构和特征”这一机制，即还要看它是如何制作出来的。而中国则仅以“具有内容生成能力”这一输出能力来进行定义，即只看它能制作出什么。

这一差异在边界案例中分化出了不同的结果。设想一个基于规则合成音乐的系统。由于成果物是声音，它可能会被纳入中国的定义中。然而，因为它并非“模仿”学习数据的结构，而是按照预设规则组合音符，所以很难纳入韩国的定义中。韩国的定义比中国的定义更窄。

中国的《暂行办法》中包含韩国所没有的独特要素。这就是内容控制义务。服务提供者应当保证遵守“社会主义核心价值观”。禁止出现煽动颠覆国家政权、恐怖主义、民族仇恨的内容。具有“舆论属性或者社会动员能力”的服务被要求进行事前安全评估。无论是在韩国《人工智能基本法》还是欧盟《人工智能法案》中，都没有这种性质的义务。

登记（备案）制度是核心执行手段。具有舆论属性或者社会动员能力的生成式AI服务在进行安全评估的同时，还必须向国家互联网信息办公室（CAC）进行算法备案，备案申报应当在开始提供服务之日起10个工作日内完成。在实务中，绝大多数公开的生成式AI服务均被按符合此要求的方式进行管理。据悉，截至2025年11月，已有611个生成式AI服务完成了备案，算法备案的总件数已超过5000件。

韩国则没有这种备案义务。在启动生成式人工智能服务时，无需向政府申报或获得许可。只存在事后自律监管和生成物标识这一透明度义务。

于2025年9月1日起施行的中国关于人工智能生成合成内容标识的办法规定了双重标识义务。必须同时附加显性标签和元数据或数字水印。甚至还制定了强制性国家技术标准（GB 45438-2025）。而韩国的生成物标识义务，则是通过《施行令》规定标识方法的大体框架，并由《指南》引导具体细节，并无强制性技术标准。

将这三个国家的路径用一句话对比，结果如下：韩国是以创造出什么（输出类型）为着眼点；欧盟是以能有多广泛的用途（通用性）为着眼点；中国则是以是否有创造能力（输出能力）为着眼点。

8. 3000件商品描述

让我们回到江南的那家电子商务公司。

该公司使用的AI系统输入商品照片并生成文本。它学习了大量的商品描述数据，通过模仿其结构和特征来生成新的描述文本。这满足了第2条第5项的三个条件，属于生成式人工智能。

如果该公司直接开发了AI模型，则属于开发经营者。如果其通过调用外部API并以此成果物向使用者提供服务，则属于使用经营者。无论是哪一种，由于均属于经营者，因此都需要承担生成物标识和事先告知的义务。

这3000件商品描述中必须加上“由AI编写”吗？在显而易见属于AI生成物的情况下，是承认例外的。然而，如果使用者无法区分人写的商品描述和AI写的商品描述，则需要进行标识。

同一位团队负责人打开了另一个屏幕。上面显示着自家购物网站的“推荐商品”列表。AI分析每位顾客的购买历史，挑选出列表并展示。这属于筛选而非生成。因此，不适用生成式人工智能的额外义务。在同一家公司里、同一天里使用的同

一款AI技术，在这一端属于生成式，而在另一端则不属于。

法律有些地方是明确的，有些地方则是不明确的。ChatGPT、Midjourney、Suno AI是明确的。推荐算法则明确排除在外。翻译引擎、搜索自动补全、AI美颜修图滤镜处于边界之上。虽然科学技术信息通信部的指南和《人工智能基本法》支持服务台的咨询正在逐行填补这些边界，但技术的发展速度快于法律的解释。生成式人工智能这一范畴的边界究竟会延伸到哪里，将随着服务的增加和案例的积累而逐渐明朗。

第四章. 第二阶段判定：我们是高影响人工智能经营者吗？

2025年秋天，一家中型人事管理公司的会议室。该公司运营着AI面试分析系统。应聘者进行视频面试时，AI会分析其语音语调、回答内容和表情变化，从而计算出适合度分数。作为客户的大企业人事团队在决定是否通过第一轮面试时会参考该分数。2026年1月22日，韩国《人工智能基本法》开始施行。该公司的法务专员翻开该法第2条第4项。第7目规定，“在招聘、贷款审查等对个人的权利、义务关系产生重大影响的判断或评估。”目光停滞了。我们的系统是否属于此列？如果属于，就必须制定风险管理方案、说明判断依据、制定保护使用者的方案、建立人员可介入的体系，并将文件保存5年。如果不属于，则没有这些义务。这一行的判定，关乎着数亿韩元的合规成本。

1. 什么是高影响人工智能

《人工智能基本法》第2条第4项予以定义：“对人的生命、身体安全及基本权利产生重大影响或有带来风险之虞，且在下列各项任意一个领域中使用的人工智能系统。”

必须同时满足两个条件。必须在法律列举的领域中使用，并且必须对生命、身体安全或基本权利产生重大影响。即使属于规定领域，如果影响不重大，也不属于高影响人工智能。即使影响重大，如果在列举的领域之外使用，也不属于高影响人工智能。

这就是两阶段判定结构。

第一阶段：领域适用性。该人工智能系统在法律第2条第4项第1目至第11目列举的十一个领域中，到底在何处被使用。

第二阶段：重大性判定。该系统在该领域中，是否对人的生命、身体安全及基本权利产生重大影响，或是有带来风险之虞。

施行令第25条也规定了这一判定的程序。经营者必须事先探讨自身的系统是否属于高影响人工智能，必要时可以向科学技术信息通信部部长申请确认。接到申请的部长基本上有30天的时间进行回复。只能延长一次，为期30天。对于经营者而言，将事先探讨的判定依据以文件形式留存下来比较安全。

2. 十一个领域

法律第2条第4项列举了从第1目到第11目的十一个领域。

第1目。依据《能源法》进行的能源供应。即在电力供应链、燃气管道控制、供热系统中使用AI的情况。如果电力供需预测模型发生故障，就会导致大范围停电。2011年9月15日韩国全境发生的轮流停电事件，就是由电力供需预测错误引起的。在AI替代这种预测的时代，该AI的失败就是基础设施的失败。

第2目。依据《饮用水管理法》进行的饮用水生产工艺。对象为在净水厂分析水质数据并决定药品投放量的AI。错误的药品投放会导致水质污染，而水质污染则会引发疾病。这一领域的范围很窄。仅限于“生产工艺”。不包括自来水管道的管理或饮用水流通。

第3目。依据《保健医疗基本法》进行的保健医疗提供及使用体系的构建与运营。包括医院信息系统、远程诊疗平台、患者分配算法、医疗资源分配系统。范围很广。涵盖了保健医疗体系的构建与运营的全过程。AI辅助急诊患者分类（triage）的系统、AI决定病床分配的系统、AI分析体检结果并建议追加检查的系统，这全部都可能落入第3目。

第4目。依据《医疗器械法》的医疗器械以及依据《数字医疗产品法》的数字医疗器械的开发与使用。AI影像判读软件、基于AI的心电图分析仪、AI诊断辅助程序皆被纳入此处。在该项中，关于数字医疗器械的部分于2026年1月24日施行。比原法施行的1月22日晚了两天。这是因为《数字医疗产品法》在同一天施行。依据《医疗器械法》的医疗器械部分则与原法一同自1月22日起适用。医疗器械经营者已经面临《医疗器械法》下的许可审批体系。《人工智能基本法》在此之

上又增加了一项高影响人工智能义务。不过，因为有放宽重复监管的条款，所以已经履行《医疗器械法》上义务的经营者，可以在相应范围内被认定为已履行基本法的义务。

第3目的保健医疗和第4目的医疗器械存在重叠的领域。一个AI影像判读系统可能同时符合这两个项目。因为其既是医疗器械，同时也被用于保健医疗体系的运营。在这种情况下，无论归属哪一方都属于高影响人工智能，所以在义务范围上，实务上的差异并不大。

第5目。依据《核设施等防护及放射能防灾对策法》进行的核物质及核设施的安全管理与运营。即在核电站控制系统、核燃料管理、辐射监测系统中应用AI的情况。核能是一个失败成本接近无穷大的领域。在该领域使用AI的经营者，已经受到《原子能安全法》体系的严格监管。

第6目。为了犯罪侦查或逮捕，对生物特征识别信息进行的分析与使用。法律将面部、指纹、虹膜、手掌静脉列为生物特征识别信息的示例。对象为侦查犯罪目的的人脸识别AI、指纹匹配AI、虹膜扫描分析AI。其用途是受限的。必须是“为了犯罪侦查或逮捕”。出于出入境管理或建筑物安保目的的人脸识别不属于第6目。不过，其有属于第7目或第9目的余地。

EU AI Act对待生物特征识别比韩国更为严格。原则上禁止将实时远程生物特征识别系统（real-time remote biometric identification）用于执法目的。只允许防止恐怖袭击或搜寻失踪儿童等极窄的例外情况。韩国法律则没有这种源头禁止规定。只是将犯罪侦查目的的生物特征识别AI分类为高影响人工智能并赋予确保安全性义务，并未禁止使用。

第7目。在招聘、贷款审查等对个人的权利、义务关系产生重大影响的判断或评估。“等”字敞开了范围。招聘和贷款审查只是示例。保险参保审查、助学贷款审查、租赁资格判断、社会保障津贴领取资格判定，这全部都可以被包含在“对个人的权利、义务关系产生重大影响的判断”中。

前面提到的AI面试分析系统很有可能属于第7目。如果该系统的适合度分数实质性地影响招聘决定，则同样满足第二阶段判定的重大性条件。但如果该分数仅作为参考资料，由人事专员独立进行判断，则存在被排除在重大性判定之外的余地。最终，AI在招聘决定中所占的分量，决定了其是否属于高影响人工智能。

第8目。依据《交通安全法》的交通工具、交通设施、交通体系的主要运行及运营。对象为自动驾驶汽车的行驶控制、航空管制辅助、铁路运行管理、信号系统优化。附有“主要运行及运营”这一限定词。由于交通信息导航App或导航仪的路径推荐不属于交通设施的“主要运行”，因此其极有可能不属于第8目。但是，控制自动驾驶汽车转向和制动的AI本身就是交通工具的主要运行。

第9目。提供公共服务所需的资格确认及决定，或费用征收等对国民产生影响的国家、地方自治团体，以及依据《有关公共机构运营的法律》第4条的公共机构等的决策。即在公务员考试评分、福利津贴领取资格判定、公用事业收费标准制定中使用AI的情况。因为主语是国家、地方自治团体和公共机构，所以民间经营者直接成为当事人的情况并不多见。但是，如果民间企业向公共机构供应AI系统，且该系统被用于第9目所指的决策中，那么作为开发经营者可能会承担高影响人工智能的义务。

第10目。依据《教育基本法》对幼儿教育、初等教育及中等教育的学生评估。对象为AI对学生的考试进行评分、评估学业成就度或诊断学习水平的系统。范围是受限的。不包括大学教育。仅适用于幼儿至高中生。非学生评估的学习辅助（AI家教、题目推荐）由于不属于评估，因此不是第10目的对象。

第11目。其他对人的生命、身体安全及基本权利保护产生重大影响，并由总统令规定的领域。截至2026年7月，尚未有由总统令追加指定的领域。该项目是开放性条款。这是为了随着技术发展或社会变化可以追加新领域而留出的空间。

3. 第二阶段判定的实际情况

施行令和高影响AI判定指南是这样运用第二阶段判定的。

第一阶段相对明确。对照我们的系统是在十一个领域中的哪一个领域中使用。能源供应、饮用水生产、保健医疗、医疗器械、核能、生物特征识别犯罪侦查、招聘或贷款审查等权利义务判断、交通、公共服务、学生评估。如果不属于上述任何一个领域，就不是高影响人工智能。如果属于相应领域，则进入第二阶段。

第二阶段比较棘手。何为“重大影响”？法律写道：“对人的生命、身体安全及基本权利产生重大影响或有带来风险之虞”。施行令第25条第2款提出了判定的考量因素。即对人的生命、身体安全及基本权利可能造成的风险的影响、重大性、频率，以及各使用领域的特殊性。

我们将分为具体的案例来看。

AI招聘面试系统。第一阶段判定。属于第7目的“招聘”。第二阶段判定。如果AI的适合度分数实质上决定了是否通过第一轮面试，这就对应聘者选择职业自由这一基本权利产生了重大影响。该系统就是高影响人工智能。相反，即使是同一个系统，如果只是参考AI的分数，由人事专员独立对所有人员进行面试并做出最终决定的结构，则很难认为AI的影响是重大的。在“人来做决定，AI只是辅助”的结构下，可能无法跨越第二阶段。这条线将是实务中频繁发生争议的地方。

AI信用评估系统。第一阶段判定。属于第7目的“贷款审查”。第二阶段判定。如果信用分数直接关系到贷款审批或利率决定，就会对个人的财产权产生重大影响。如果是原先受《信用信息法》适用的信用评估公司，依据放宽重复监管条款，在履行了《信用信息法》规定的义务范围内，可被认定为履行了基本法的责任。

AI学生评估系统。第一阶段判定。属于第10目的“学生评估”。第二阶段判定。如果该AI负责期中考试或期末考试的评分，且该评分结果反映在内审成绩中，就会对学生受教育的权利和升学机会产生重大影响。这是高影响人工智能。如果AI仅仅止步于推荐练习题或展示学习进度，则由于不属于“评估”而不适用第10目，自然也没有必要进入第二阶段。

4. 是“影响”还是“风险”

韩国法制研究院研究委员郑元俊在2025年9月的报告中指出了该法在用语上的混乱。

我们再来看一下法律第2条第4项的定义：“对人的生命、身体安全及基本权利产生重大影响或有带来风险之虞的人工智能系统。”在一个句子中，“影响”和“风险”同时出现。在“产生重大影响”之后紧跟着“有带来风险之虞”。

“影响（impact）”和“风险（risk）”的概念是不同的。影响是指已经发生的结果或预计发生的结果的程度。风险则是发生有害结果的概率与严重性的乘积。韩国法律的名字是高影响。但在定义条款中也包含了风险一词。

郑元俊研究委员的批评是这样的。在法律中并行使用“影响”和“风险”，导致经营者在应该以何为基准进行判断上产生混乱。并且，在“人的生命、身体安全及基本权利”这一短语中，“及（and）”将安全和基本权利并列连接，那么，是即使不影响安全但影响了基本权利也属于此列，还是必须同时对安全和基本权利双方产生影响，在解释上出现了分歧。

从经营者的立场来说是这样的。AI招聘面试系统不会对人的生命或身体安全产生影响。但会对基本权利（选择职业的自由）产生影响。如果把“及”读作“和（and）”，就会得出既然对生命和身体安全没有影响，就不属于高影响人工智能的解释。如果把“及”读作“或（or）”，就会得出即使只对基本权利产生影响，也属于高影响人工智能的解释。科学技术信息通信部的指南正朝着后者的方向运用，即即使只对基本权利产生影响的情况，也可能属于高影响人工智能。

回顾国会科学技术信息通信委员会的审议过程，在法律名称中选择“影响（impact）”而非“风险（risk）”，是为了避免给技术打上负面烙印的一种政治妥协。正如李海旻议员的发言所示，人们认为影响这个词比风险更加中立且包容。然而，由于在法律的定义条款中同时使用了影响和风险，使得名称的中立性和定义的明确性之间产生了紧张关系。

欧盟不存在这个问题。他们从一开始就使用了“风险（risk）”一词，并建立了四个层级的体系。不可接受的风险（unacceptable risk）、高风险（high-risk）、有限的风险（limited risk）、最低的风险（minimal risk）。用语和体系是一致的。

5. 欧盟的高风险AI：双重路径与豁免

EU AI Act通过两条路径对高风险AI进行分类。

第一条路径。产品安全路径（第6条第1款）。根据欧盟协调立法（附件I中列举的医疗器械、机械类、玩具、航空器、汽车等产品相关指令），如果AI被用作必须接受第三方合格评定的产品的安全组成部分，则属于高风险。进入医疗器械的AI诊断软件就是典型的例子。在这一路径中，原有的产品安全体系和AI监管产生了重叠。

第二条路径。附件III路径（第6条第2款）。对象为在附件III列举的领域中使用的独立型AI系统。生物特征识别、核心基础设施、教育和职业培训、就业和人员管理、获得必要的公共和私人服务、执法、移民、庇护和边境管理，以及司法和民主程序。这与韩国法律的十一大领域在很大程度上重合，但也存在差异。欧盟明文规定了移民和边境管理、司法和民主程序，而韩国却没有列举这些领域。另外，韩国将能源供应、饮用水生产和核能单独分列成项，而欧盟则将这些统归为“核心基础设施”这一项。

欧盟拥有一项韩国没有的机制。那就是豁免机制（第6条第3款）。即使是在附件III列举的领域中的AI系统，如果不给人的健康、安全或基本权利带来具有重大意义的风险（significant risk），就不被分类为高风险。执行狭义的程序性任务、旨在改善之前由人做出的决定、或在决策中提供帮助但不替代人的评估的情况属于豁免事由。不过，针对自然人的画像（profiling）始终是高风险。不适用任何豁免。

韩国法律中没有这一豁免机制。只要属于十一个领域并且具有重大影响，就是高影响人工智能。在法律条文中并没有像欧盟那样，存在“该系统虽然在附件III的领域中使用，但实质性风险较低，因此不是高风险”这样的退出路径。虽然在第二阶段判定中，可以通过对“重大影响”的解释得出类似的结论，但并不像欧盟那样列举了明文的豁免条件。

韩国和欧盟的另一个差异在于义务的强度。欧盟对高风险AI施加了强制性合格评定（conformity assessment）。在上市前必须编制技术文件、具备质量管理体系、发布符合性声明，并贴上CE标志。有时还需要第三方机构的评估。如果不合格，最高将被处以1500万欧元或营业额3%的罚款。

韩国高影响人工智能经营者义务的性质有所不同。法律第34条第1款在一个条款中规定了五项。必须制定并运营风险管理方案。必须制定并施行关于判断的主要基准和学习用数据概况等的说明方案。必须制定并运营使用者保护方案。必须具备人进行管理和监督的体系。必须编写履行依据文件并保存5年。在此基础上，施行令第27条第1款增加了公布义务。要求在办事处、营业场所或互联网主页等处公布风险管理方案的主要内容、说明方案的主要内容、使用者保护方案以及管理、监督者的姓名和联系方式，但属于商业机密的事项可以剔除。义务列表虽然很实际，但并没有像第三方合格评定那样的预先关卡。这也不是违反了该责任就立即被处以行政罚款的结构。只有在接到政府的纠正命令仍不履行时，才会被处以3000万韩元以下的行政罚款。政府表示了将在施行后设定至少1年缓冲期的方针。

6. 中国：不进行风险分类而划分领域

中国根本不使用“高风险AI”或“高影响人工智能”这一范畴。

不像欧盟那样存在以风险等级横跨所有AI的水平分类体系。只是在个别规定内设有分类、分级管理原则。取而代之的是按技术类型划分了规定。对于算法推荐，有《算法推荐管理规定》（2022年3月施行）；对于深度伪造，有《深度合成

管理规定》（2023年1月施行）；对于生成式人工智能，有《生成式人工智能服务管理暂行办法》（2023年8月施行）。2025年追加了《人工智能生成内容标识办法》，而在2026年7月则施行了《拟人化AI交互服务暂行办法》，在此之前的5月，还发布了《关于AI智能体规范的实施意见》。

这与韩国用一部法律涵盖所有AI，并在其中划分为高影响人工智能和普通人工智能的做法恰好相反。中国是每当一种技术出现时，就出台适合该技术的规定。每项规定只定义自己管辖内的技术，并赋予适合该技术的义务。

在中国，勉强接近风险分类的，是“舆论属性或社会动员能力”这一基准。《生成式人工智能服务管理暂行办法》要求具有此类属性的服务进行安全评估。具有“社会动员能力”的AI服务，必须事先向国家互联网信息办公室（CAC）申请安全评估，只有通过了该评估，才能提供服务。这与韩国的高影响人工智能或欧盟的高风险是不同的基准。其判定的轴心不是对侵害基本权利的担忧，而是维持社会稳定。

中国正在筹备包容性的人工智能法。国务院在2026年5月的年度立法工作计划中表示，“加快推进促进人工智能健康发展的综合性立法”。学者建议稿（2024年3月，由中国政法大学等学界共同提出）中包含了基于AI负面清单的许可管理制度，而学界预测全国人民代表大会常务委员会将在2027年底左右通过，并于2028年施行。一旦制定了该综合法，看来中国也会以某种形式引入风险等级分类。

纵观这三个国家，情况是这样的。

韩国设立了“高影响”这一单一范畴，列举了十一个领域，并通过两阶段判定来甄别是否属于此列。没有禁止的范畴。

欧盟建立了四个层级的风险分类，设定了被禁止的惯例，在对高风险AI设置双重路径的同时，还配备了豁免机制。其体系非常精细。

中国没有采用欧盟式的风险等级分类，而是按技术类型划分规定，并将内容控制与事先备案结合起来。

7. 经营者的判定程序

在此整理经营者自行判定是否属于高影响人工智能的程序。

步骤一。我们的系统是否属于人工智能系统？是否满足法律第2条第2项的条件（自主性、适应性、推理）？仅仅遵循既定规则的自动化被解释为无法满足这一条件。

步骤二。我们是否为人工智能经营者？是否开发并提供该系统，或是获得提供后开发成服务进行提供？如果只是获得提供并加以使用，那就是使用者。

步骤三。是否在十一个领域中的至少一个领域中使用？如果不属于这些领域，那就不是高影响人工智能。

步骤四。是否对生命、身体安全、基本权利产生重大影响？要探究其影响和风险的重大性、发生频率，以及各使用领域的特殊性。

步骤五。将判定依据以文件形式留存。无论是否属于高影响人工智能，将判定过程和依据记录下来是比较安全的做法。如果难以自行判定，可以依照施行令第25条的程序向科学技术信息通信部部长申请确认，接到申请的部长将在30天（延长一次最多为60天）内回复结果。如果对回复结果有异议，可以在收到回复之日起10天内申请重新确认。

如果属于高影响人工智能，随之而来的便是法律第34条第1款在一个条款中以五项规定的义务。制定并运营风险管理方案。制定并施行针对使用者的判断依据说明方案。制定并运营使用者保护方案。人员的管理、监督。以及编写履行文件并保存5年。

8. 处于边界上的系统

有一些系统并不明确是否属于高影响人工智能。

基于AI的保险审查系统。由AI分析保险参保申请，并决定批准、拒绝或保险费。这可能属于第7目的“贷款审查等对个人的权利、义务关系产生重大影响的判断”。关键在于“等”字是否包含保险审查。拒绝参保会直接影响个人的经济权利。有被分类为高影响人工智能的可能性。

基于AI的交通信息服务。分析实时交通数据并推荐最优路径。第8目写的是“交通工具、交通设施、交通体系的主要运行及运营”。路径推荐并不是交通体系的主要运行或运营。它是向驾驶员提供便利的辅助服务。极有可能不是高影响人工智能。

基于AI的公寓管理系统。优化能源消耗以管理集中供暖和电力。第1目写的是“依据《能源法》进行的能源供应”。公寓内的能源管理是否属于依据《能源法》的能源供应？《能源法》中的能源供应是指电力公司或燃气公司向消费者供应能源的行为。公寓管理系统是分配所接收到的能源，而非供应能源。极有可能不属于此类。

基于AI的国家奖学金领取资格判定系统。属于第9目的“国家、地方自治团体、公共机构的资格审查、确认、判定”。在第二阶段判定中，也会直接影响学生的受教育机会。这是高影响人工智能。

边界案例积累得越多，第二阶段判定中“重大影响”一词所带有的不确定性就会显露出来。这种不确定性不是缺陷，而是设计本身。《基本法》不可能事先对所有案例进行分类，因此留出了判断的余地，其结构就是通过指南、权威解释和案例积累来加以填补。不过，对于经营者而言，这种不确定性就是成本。如果判断出错就是违反义务，如果过度遵守则是徒增的支出。

9. 纵观三个国家的体系

让我们回到人事管理公司的法务专员这里。

这一AI面试分析系统在韩国法律中极有可能属于高影响人工智能。第一阶段，第7目的招聘领域。第二阶段，对应聘者选择职业自由产生重大影响。因此，必须制定风险管理方案、建立说明判断依据的体系、制定使用者保护方案、配备人员介入机制，并将文件保存5年。

如果将同样的系统推向欧洲市场，由于属于EU AI Act附件III的就业领域，除非包含在豁免事由中，否则将被分类为高风险AI。这就需要接受合格评定、编制技术文件，并附上CE标志。不过，由于2026年的数字综合法案修订，附件III高风险AI义务的适用被推迟到了2027年12月2日。比较一下制裁上限，单是1500万欧元，就已经是韩国行政罚款上限3000万韩元的700多倍，如果是适用营业额3%基准的大企业，这个差距将扩大到数千倍。

如果在中国提供相同系统的服务，因为根本没有高风险这一分类，所以必须按个别规定逐一确认是否存在可适用的规定。因为不是生成式人工智能，所以不适用《生成式人工智能服务管理暂行办法》。如果使用了算法推荐，则可能适用《算法推荐管理规定》的备案义务。

在高影响人工智能这一行定义背后，交织着经营者的成本与使用者的保护。虽然法律的名称是高影响，但在条款中也同时存在着风险一词。第二阶段判定的重大性基准仍处于用案例填补的过程中。随着科学技术信息通信部部长的确认程序、指南的运用、《人工智能基本法》援助服务台的咨询案例逐一积累，边界的轮廓将会逐渐清晰。在当前这个时间点，经营者能做的事情是：履行五阶段判定程序、将判定依据以文件形式留存下来，并且如果处于边界上，则采取保守的判定。比起接到确认请求后才表明自己不是高影响人工智能，将其视为高影响人工智能并履行义务后发现只是防卫过当，后者的风险要小得多。

第5章. 第3步判断：我们是高性能、最前沿人工智能经营者吗？

2023年，AI研究机构Epoch AI估算了GPT-4的训练计算量。约为2乘以10的25次方FLOP。这是数千台GPU连续运转数月的结果。据估算，训练投入的费用为1亿美元，折合韩元约1400亿韩元。即使是这个GPT-4，也可能达不到韩国《人工智能基本法》施行令所规定的高性能人工智能标准。施行令第24条规定的计算量门槛是累计训练计算量10的26次方FLOP。这是GPT-4估算值的五倍。

1. 10的26次方这一门槛

高性能人工智能这一概念并未出现在《人工智能基本法》第2条的十二项定义中。基本法第32条宣布，将对达到总统令所定标准及以上的人工智能系统赋予确保安全性义务，而施行令第24条将该标准予以了具体化。标准共有三项，且必须全部满足。其一，用于训练的累计计算量在10的26次方FLOP以上；其二，应用当前被运用的人工智能技术中最前沿的技术进行构建和运营；其三，其风险度存在对人的生命、身体安全及基本权利产生广泛而重大影响的隐患。即使超过了计算量门槛，如果未能满足其余两项要件，也无法被纳入该标准。

FLOP是浮点运算（floating-point operation）的缩写。这是计算机进行一次带小数点数字计算的单位。要训练AI模型，需要经过数十亿次反复读取数据、计算误差并调整权重的过程。将所有这些计算加起来，就是累计训练计算量。

10的26次方是1后面带26个0的数字。要产生这个数字，需要数万台最新GPU连续运转数月。加上电力、冷却和数据中心建设，这是数千亿韩元规模的投资。全球能够承担这种规模的组织不到十家。

施行令第24条第2款授权科学技术信息通信部长官以公告形式规定累计计算量的具体计算方式。科学技术信息通信部于2026年初针对有关履行确保安全性义务方法的公告制定案进行了行政预告。公告能够规定的仅限于计算量的计算方法。

如果要更改以何种人工智能为对象这一标准本身，就必须修改施行令。

2. 哪些模型会越过这一门槛？

在 10^{26} 这个数字面前，需要考量一下实际模型所处的位置。

问题在于几乎无人知晓准确的训练计算量。OpenAI、Google DeepMind、Anthropic、Meta等任何公司都未曾正式公开过自家模型的训练计算量。这都是由Epoch AI等独立研究机构结合公开信息、技术论文及业界证言推算出来的。这是估算值，并非官方数据。

根据Epoch AI的估算，GPT-4的训练计算量约为2乘以 10^{25} 次方FLOP。Google的Gemini 1.0 Ultra估算约为5乘以 10^{25} 次方。Meta的Llama 3.1 405B约为3.8乘以 10^{25} 次方。全部处于 10^{25} 次方量级，均未达到 10^{26} 次方。

不过这一格局已在发生变化。Epoch AI估算，2024年以后训练的部分前沿模型（如xAI的Grok 3）的训练计算量已超过 10^{26} 次方。但这终究只是独立机构的推算，并没有开发商官方确认的数据。

韩国的门槛设定在一个连2023年世界最高水平模型也无法跨越的高度。这并非偶然。

3. 韩国与欧盟：十倍的差距

EU AI Act（《欧盟人工智能法案》）使用 10^{25} 次方FLOP作为甄别通用人工智能（GPAI）模型中是否存在系统性风险的标准。这与韩国的 10^{26} 次方相差十倍。不同的不仅是数字。欧盟的 10^{25} 次方是一种设定，只要超过该界限，便初步推定存在系统性风险；而韩国的 10^{26} 次方则是一个组合标准，只有在同时满足应用最前沿技术及存在重大风险隐患这两个要件时，才会成为适用对象。

十倍不仅仅是数字上的差距，更是政策上的差异。监管对象的范围也随之改变。

如果适用欧盟10的25次方标准，GPT-4（约2乘以10的25次方）、Gemini Ultra（约5乘以10的25次方）以及Llama 3.1 405B（约3.8乘以10的25次方）均超过了该标准。开发这些模型的OpenAI、Google、Meta在欧盟法律下被归类为具有系统性风险的GPAI提供商，将承担透明度义务、技术文档化以及系统性风险缓解措施等额外义务。

如果适用韩国10的26次方标准，同样的模型就会降至标准线以下。GPT-4仅为标准的五分之一；Gemini Ultra为一半。在韩国法律下，这些模型均不属于高性能人工智能。

锁定监管对象的幅度有所不同。欧盟广泛界定了潜在风险，并宽泛地设定了监管对象；韩国则采取了仅锁定那些确实可能产生广泛影响的最顶层模型的态度。

意图很明确。政府的判断是，在国内AI产业尚处成长阶段时，应通过缩小监管对象来减轻经营者的负担。10的26次方这一高门槛，实际上是一个让现阶段几乎没有韩国国内经营者需要履行该法律中高性能AI义务的机制。

对这一判断也有反对意见。学术界有观点指出，韩国的标准比欧盟高出一个数量级，可能会产生监管盲区。由于训练计算量每年都在急剧增加，今天的高门槛明天可能就会变成普通水平。如果标准固定在法律中，那将是一个严重问题；但韩国法律已将标准授权给施行令。只要修改总统令，就能降低门槛。

中国并未采用基于计算量的分类方法。由于采取的是通过各领域的个别规定来管控技术的方式，所以根本不存在根据训练计算量来划分监管水平的结构。不过，在讨论制定综合性人工智能法案的过程中，也有提及对参数超过1000亿的模型赋予义务性安全影响评估的方案。从以参数数量而非计算量（FLOP）为标尺这一点来看，这与韩国及欧盟的做法是不同的途径。

4. 对高性能AI经营者有什么要求？

对于开发并提供满足施行令第24条全部三项要件的人工智能经营者，将被赋予确保安全性的义务。基本法第32条为该义务奠定了法律依据，而施行令第24条则限定了适用对象。

义务的内容贯穿模型的整个生命周期。

必须识别风险。必须事先掌握模型可能产生哪些有害结果，以及存在哪些被滥用的可能性。必须对识别出的风险进行评估。需要衡量其发生的可能性及损害规模。然后必须根据评估结果采取缓解措施。技术上的补充、使用限制以及监控体系便是此类方法。必须将这一过程和结果制作成文件并予以提交。

必须构建能够监控和应对人工智能相关安全事故的风险管理体系。这项义务从模型训练那一刻起，一直延续到服务终止之时。

这与高影响AI经营者的义务存在重叠区间。根据第34条第1款，高影响AI经营者负有制定风险管理方案、解释判断依据、保护使用者、实施人为管理与监督、以及撰写文件并保存五年这五项措施义务。而对于高性能AI经营者，在此基础之上还增加了确保安全性的义务。如果超过10的26次方FLOP的模型被用于招聘决策或贷款审查等高影响领域，就必须同时履行这两方面的义务。

EU AI Act对存在系统性风险的GPAI模型赋予了更为具体的义务。不仅需要执行模型评估和红队测试，还必须追踪并缓解系统性风险，并将严重事件（serious incidents）报告给欧盟AI办公室。至于编写并维护技术文件以及公开关于训练数据的足够详细的摘要等义务，则是无论是否存在系统性风险，所有GPAI提供商都必须共同承担的。2025年8月经欧盟委员会评估为妥当的GPAI行为准则，由透明度、版权和安全这三大支柱构成。它要求针对四种类型的系统性风险，即CBRN（化学、生物、放射性、核）威胁能力、失控、协助网络攻击以及有害的操纵和欺骗，采取具体的应对措施。

相比之下，韩国法律中确保安全性的义务仅停留在轮廓层面。虽然具备了风险识别与评估、缓解以及文档化这一框架，但像红队测试或系统性风险报告这样的详细履行方法，仍处于在公告和指南中逐步完善的过程之中。

5. 韩国国内AI企业与10的26次方之间的距离

Naver的HyperCLOVA X。三星电子的Gauss。LG AI Research的EXAONE。KT的Mi:dm。Kakao的Kanana。

截至2026年，韩国国内有多家企业在开发自有的大型语言模型。尽管这些模型的训练计算量从未被官方公布，但根据业界的普遍估算，它们大都停留在10的23次方到10的24次方FLOP范围内。这与10的26次方有着百倍到千倍的差距。

GPU的供需是核心瓶颈。要长期拥有数万个NVIDIA的A100或H100 GPU并进行大规模训练，需要数万亿韩元的投资。OpenAI获得了Microsoft数十万亿韩元规模的投资；Google运营着自有的TPU基础设施；Meta则在其数据中心部署了数十万个GPU。韩国国内企业与这一规模完全不在一个层面上。2025年9月，国家人工智能战略委员会审议并表决了有关推进国家AI计算中心方案，这也是政府层面面对此差距做出的回应。

在可公开资料确认的范围内，现阶段似乎没有哪家韩国国内AI经营者满足施行令第24条的高性能人工智能标准。尽管法律已经施行，但该条款所瞄准的适用对象并非韩国企业。

可能成为适用对象的经营者是全球大型科技公司。如果由OpenAI、Google、Meta、Anthropic开发的前沿模型在满足施行令第24条三项要件的情况下向韩国市场提供，那么就将适用基本法。根据第4条第1款的域外适用条款，即便是在国外的行为，只要对韩国国内市场或使用者产生了影响，就必须遵守该法。如果相关海外经营者符合韩国境内代理人指定标准（即上一年度销售额达到1万亿韩元以上、人工智能服务部门上一年度销售额达到100亿韩元以上、截至上一年度末的前三个月内韩国国内使用者数量日均100万人以上、有根据该法被处以行政罚

款的履历，这四项标准中的任意一项），就必须指定韩国境内代理人，并通过其处理诸如提交确保安全性义务履行结果等业务。

仅就高性能AI而言，韩国法律瞄准的并非韩国企业，而是海外企业。韩国国内企业尚未达到该标准，而且在短期内达到的可能性也很低。

6. 使用API的经营者的位置

假设一家韩国国内初创企业调用OpenAI的GPT-5 API来开发法律文书审查服务。如果GPT-5是超过10的26次方FLOP的高性能人工智能，那么这家初创企业也会成为高性能人工智能经营者吗？

不是。

这家初创企业并未开发GPT-5，而是在使用它。根据第2条第7项b目，它属于使用经营者。高性能人工智能中确保安全性义务归属于开发并提供该人工智能的经营者，在这里也就是OpenAI。调用API的行为与训练计算量无关。在调用API之时，模型的训练早已结束。

这家初创企业是否要承担高影响人工智能的义务，则是另一个问题。如果法律文书审查属于高影响领域（第2条第4项g目，即对个人权利和义务产生重大影响的判断），那么这家初创企业作为使用经营者，就必须承担与高影响AI相关的义务。它必须制定风险管理方案、解释判断依据，并具备允许人干预的体系。这是基于应用领域的义务，与10的26次方FLOP无关。

不直接开发大型语言模型而只是通过API予以使用的经营者，并非高性能人工智能经营者。但是，根据其在何种领域以何种方式使用该模型，它可能会承担高影响AI经营者的义务。高性能与高影响是不同的轴线。计算量的轴线与应用领域的轴线是分开运行的。

对于这种划分并不能完全放心，这是有原因的。高性能模型的风险并不仅仅停留在模型内部。根据将该模型置于何种脉络之中，风险的形态也会发生改变。如

果通过API使用高性能模型的使用经营者无法窥见模型内部的运行原理，那么在识别和管理风险时就会产生局限性。在开发经营者位于海外、而仅形式上指定了韩国境内代理人的情况下，能否进行有效的管理和监督，这一疑问依然存在。

7. 计算FLOP的难度

要使10的26次方这一标准发挥作用，就必须能够测量经营者的累计训练计算量。测量并不容易。

训练计算量的公开是一大障碍。在OpenAI、Google、Anthropic和Meta中，没有一家公司正式公布过自家模型确切的训练FLOP。这均由Epoch AI等独立研究机构通过结合公开论文、访谈及业界证言进行反向推算。这种将模型参数数量、训练代币数量、GPU种类与数量以及训练周期综合起来进行计算的方式，一旦前提改变，结果就会截然不同。对于同一模型，甚至会出现不同研究者得出不同推算值的情况。

训练的范围也很模糊。是仅计算用于预训练（pre-training）的计算量，还是连同后训练（post-training）及基于人类反馈的强化学习（RLHF）一并包含在内？用于微调（fine-tuning）的计算量是否也要加上？施行令中写的是“用于训练的累计计算量”，而将具体计算方式留给了公告。进行行政预告的公告制定案似乎包含着广泛反映与训练相关的计算方向，但在实际业务中，预训练与后训练的界限到底画在哪里，仍处于不断完善的过程之中。

测量单位本身也不明确。FLOP是理论运算次数。在实际硬件中的运算效率会因GPU的种类、并行处理方式以及精度（是16位还是32位浮点数）的设定不同而发生改变。关于如何处理理论FLOP与实测FLOP之间的差距，目前也尚未达成共识。

即便经营者自发报告自家模型的计算量，外部也很难对该数值进行验证。在现实中，要求监管机构直接去核查训练过程中的GPU日志、用电记录以及数据中心运转记录是不太容易的。这就会变成一种监管机构依赖于经营者自我报告的结构。

这一困难并非仅仅是韩国面临的课题。在EU AI Act中，关于GPAI模型训练计算量的计算，同样的讨论也一直在持续。虽然GPAI行为准则要求在技术文档中包含训练计算量，但尚未能统一其计算方法。除了10的25次方FLOP这一数值标准外，欧盟还试图通过并列“同等水平的能力（equivalent capability）”这一具有定性特征的标准，来弥补数值标准带来的局限性。

韩国施行令将应用最前沿技术及存在重大风险隐患这一具有定性特征的要件与计算量要件结合在一起，这也被解读为是对计算量这一单一指标局限性的防备。不过，公告能够触及的仅仅是计算量的计算方式。如果想要引入诸如参数数量或使用者数量等新指标作为对象标准，就必须经过施行令的修改。

8. 空箱子

站在10的26次方这一数字之上的经营者目前是极少数。韩国国内并没有相应的企业。在全球大型科技公司的最新模型中，虽然有独立推算认为已经越过这条线的模型，但并没有开发商官方确认的数据。设定这一标准的意图是为韩国AI产业留出喘息的空间，且就韩国国内企业而言，该意图已经达成。

然而计算量在逐年上升。根据Epoch AI的追踪，前沿模型的训练计算量大约每六个月就会翻一番。按照这一趋势，10的26次方在不久的将来就会成为前沿模型的标准。今天的例外将成为明天的寻常。

修改施行令便能降低门槛。计算量的计算方式也可以通过公告进行修整。但是，更改对象标准本身并非公告而是修改施行令的事项。修改施行令通常需要40天以上的立法预告、监管审查以及国务会议。在GPT-4问世仅两年时间，前沿模型的训练计算量便成倍增长的市场中，监管机构能否跟上这般速度，仍有待观察。

对于韩国国内企业而言，高性能人工智能这一范畴目前还是一个空箱子。这个箱子被填满之时，并非韩国AI企业的模型达到10的26次方的那一天；而是跨越了计算量门槛并满足其余两项要件的全球大型科技公司模型登陆韩国市场的那一天。那一天，空箱子里将会装进法律的适用对象。

第6章. 第4阶段判断：我们是需要指定韩国境内代理人的海外经营者吗

2026年春，首尔江南的一间共享办公室会议室里。从美国总部飞来的人工智能企业亚太区法务副总裁，与韩国律所的合伙人律师相对而坐。屏幕上显示着韩国《人工智能基本法》第36条。“在韩国境内没有住所或营业场所的人工智能经营者，若使用者人数、销售额等符合总统令规定的标准，应当以书面形式指定韩国境内代理人，并向科学技术信息通信部长官申报。”副总裁问道：我们在韩国设立了分公司，这样就可以了吗？律师摇了摇头：有分公司与指定法律所要求的韩国境内代理人是两码事。

1. 法律条文的内容

第36条由三款组成。

第1款规定了义务的产生条件和代理人的角色。在韩国境内没有住所或营业场所的人工智能经营者，若符合施行令规定的标准，应当以书面形式指定代理法律列举事项的人，并向科学技术信息通信部长官申报。

韩国境内代理人代理的事项有三项。提交关于高性能人工智能确保安全性义务履行情况的结果（第32条第2款）。申请确认是否属于高影响人工智能（第33条第1款）。支持履行与高影响人工智能相关的确保安全性和信赖性措施（第34条第1款各项）。第三项角色中包括了对确保安全性文档的最新性和准确性进行检查。

第2款规定了资格。韩国境内代理人必须是在韩国境内有住所或营业场所的人。

第3款是归属规定。若韩国境内代理人就上述三项职责违反了本法，被视为实施该行为的主体并非代理人，而是指定该代理人的境外经营者。代理人的违法行为后果归属于本人。并非设立了代理人，责任就会转移给代理人。

2. 四项标准：只要满足其中一项即产生义务

施行令第29条规定了这一标准。

第1项。上一年度销售额达到1万亿韩元以上的人工智能经营者。如果是法人，则以合规的上一营业年度为准。该销售额不仅指人工智能业务的销售额，而是该经营者的总销售额。由于是境外经营者，这实际上相当于全球销售额。外币销售额按照上一年度的平均汇率折算为韩元（施行令第29条第2款）。

第2项。上一年度人工智能服务部门销售额达到100亿韩元以上的人工智能经营者。第1项考察的是公司整体销售额，而第2项则仅考察人工智能服务部门的销售额。

第3项。以上一年度末为准，该经营者的人工智能产品及服务在紧邻的前3个月内，在韩国境内的日均使用者人数达到100万人以上的人工智能经营者。

第4项。曾被处以第43条第1款第3项规定之行政罚款的人工智能经营者。第43条第1款第3项是指因未履行中止命令或纠正命令而被处以行政罚款的情形。该条款意味着，一旦境外经营者受到过行政处罚，此后无论其销售额或使用者人数如何，均须承担指定韩国境内代理人的义务。

满足上述四项标准中的任意一项即产生义务。“符合下列各项之一的”即为依据。

3. 全球科技巨头是否符合标准

有报道称，截至2024年中期，OpenAI的年度经常性收入（ARR）已超过34亿美元。虽然ARR与法律所称的上一年度销售额并非完全相同的概念，但折合韩元已超过4万亿韩元，因此认为其满足第1项标准（总销售额1万亿韩元）并无不妥。ChatGPT在韩国的月访问量达数千万次。其也极有可能满足第3项标准（韩国境内日均使用者100万人）。

Google在2024年Alphabet的总销售额超过了3500亿美元。满足第1项标准。由于其在韩国提供Gemini、Google Cloud AI服务，因此也有可能满足第3项标准。

Google在韩国设有名为“谷歌韩国有限责任公司”的法人实体。这可能使其不符合“在韩国境内没有住所或营业场所的人工智能经营者”这一前提条件。问题在这里变得复杂起来。

Meta在2024年的总销售额为1645亿美元。满足第1项标准。其正在提供韩语版的Meta AI服务。Meta在韩国设有名为“Facebook Korea”的法人实体。

据估计，Anthropic在2024年的销售额约为10亿美元左右。按上一年度平均汇率折算，很有可能超过1万亿韩元，因此可能符合第1项标准。由于Claude的韩国使用者规模未公开，因此难以判断其是否满足第3项标准。Anthropic于2026年6月开设了首尔办公室，并任命了韩国代表理事。如果该首尔办公室属于第36条所述的营业场所，则将不再符合“在韩国境内没有住所或营业场所的人工智能经营者”的前提条件。该事项的结论可能取决于人工智能服务合同的签约主体是哪家法人。

xAI、Mistral AI等后起之秀可能达不到总销售额1万亿韩元标准（第1项）。然而，人工智能服务部门销售额100亿韩元标准（第2项）的门槛要低得多，这些企业也可能已经符合该标准。切记，只要满足四项标准中的任意一项就会产生义务。

4. 设有韩国分公司就不需要指定韩国境内代理人了吗

第36条第1款的适用对象是“在韩国境内没有住所或营业场所”的经营者。如果在韩国境内设有法人或营业场所，则根本不属于该条款的适用对象。因此，没有义务另行指定韩国境内代理人。

但有需要注意的地方。

即使设有韩国分公司，该分公司是否为人工智能服务的提供主体才是关键。谷歌韩国在韩国拥有法人实体，但根据向韩国使用者提供Google的人工智能服务的合同当事人是谷歌韩国还是美国总部Google LLC，答案可能会有所不同。如果服务条款的合同当事人是美国法人，而韩国法人仅执行营销或销售支持，那么提供

人工智能服务的人工智能经营者就是该美国法人。接下来的问题是，该美国法人是否可被视为在韩国境内没有住所或营业场所，以及韩国子公司的住所是否属于母公司的营业场所。

这一解释尚未最终确定。随着科学技术信息通信部的有权解释或《人工智能基本法》支持咨询台的咨询案例不断积累，标准将会更加具体化。在业务实践中，即使设有韩国法人，也应仔细检查人工智能服务的合同当事人和提供主体，若存在不确定性，指定韩国境内代理人是更为安全的抉择。

5. 与欧盟授权代表的比较

《欧盟人工智能法案》第22条要求设立在欧盟境外的高风险人工智能系统提供商（provider）在将该系统投放到欧盟市场之前，须通过书面授权书指定设立在欧盟境内的授权代表（authorised representative）。对于通用人工智能模型的境外提供商，第54条另行规定了具有相同主旨的指定授权代表的义务。

将韩国与欧盟的制度放在一起对比，两者结构非常相似。都是要求境外经营者设立境内联络窗口。都是通过书面委托设立代理人。然而，韩国还制定了将代理人的违法行为视为本人行为的归属规定（第36条第3款）以及事先申报程序，而欧盟的结构则是当主管机关要求时出示授权书。两者的范围和深度也存在差异。

欧盟的授权代表必须保管符合性声明副本和技术文档，并在主管机关要求时予以提供，响应市场监管机构的请求，并配合纠正措施。当监管当局无法联系到境外经营者时，授权代表实际上代替该经营者步入欧盟的监管体系中。

韩国境内代理人的职责仅限于三项：提交高性能人工智能安全性履行结果、申请确认高影响人工智能、以及支持履行 high-impact AI 确保安全性措施。透明度义务（第31条）或事先告知等一般性义务并不在韩国境内代理人的法定代理范围之内。

适用标准也有所不同。在欧盟，只要是将高风险人工智能系统或通用人工智能模型投放欧盟市场的境外经营者，无论其销售额或使用者规模如何，都会产生指定授权代表的义务。而韩国则设定了总销售额1万亿韩元以上、人工智能业务销售额100亿韩元以上、日均使用者100万人以上等量化标准。韩国法律豁免了小型境外经营者的义务，仅对大型经营者强加了该义务。这是一种将比例原则通过量化门槛予以实现的制度设计。

违规时处罚的差异是巨大的。根据《欧盟人工智能法案》，未指定授权代表的最大罚款额为1500万欧元或全球营业额的3%，以较高者为准。而在韩国《人工智能基本法》中，未指定韩国境内代理人的行政罚款上限为3000万韩元。对于全球科技巨头而言，3000万韩元甚至算不上是一个影响经营决策的数字。

6. 中国：在无域外适用条款下实施监管的法律

中国没有关于人工智能的通用域外适用宣告条款。

《生成式人工智能服务管理暂行办法》（2023年8月施行）第2条仅规定“利用生成式人工智能产品向中华人民共和国境内公众提供生成式人工智能服务，适用本办法”。并没有像韩国《人工智能基本法》第4条第1款那样宣告“即使是在境内实施的行为，若对境内市场或使用者产生影响，亦适用本法”。

没有通用宣告并不意味着不存在针对源自境外之服务的规定。该办法第20条规定，如果从境外向中华人民共和国境内提供的生成式人工智能服务不符合法律法规规定，国家网信部门应当通知有关机构采取技术措施等必要措施。中国通过该条款和技术性拦截来解决管辖权问题。防火长城物理性地阻止了对未备案的境外人工智能服务的访问。若要在中国境内提供服务，必须完成算法备案并通过安全评估。如果未对服务进行备案，访问本身就会被拦截。这是一种用技术基础设施划定法律边界，以代替通过法律明文宣告域外适用的方式。

中国也没有对应指定韩国境内代理人义务的制度。境外经营者若要在中国提供人工智能服务，设立中国法人或与中国当地合作伙伴合资，这是事实上存在的惯

例。这并非只设立一个代理人就能解决的结构。

韩国通过法律条文宣告域外适用，并要求将韩国境内代理人作为境内联络窗口。欧盟则在更广泛的范围内要求授权代表，并以更高的罚款作为后盾。中国不设立通用的域外适用宣告，但设立了技术拦截和事前备案等物理屏障。三个国家针对同一个问题，拿起了不同的工具。

7. 3000万韩元行政罚款的分量

若未指定韩国境内代理人，将根据第43条第1款第2项被处以3000万韩元以下的行政罚款。

我们来审视一下这一金额的意义。对于年销售额达数十万亿韩元的全球科技巨头而言，3000万韩元与其说是制裁金，倒不如说更接近于手续费。与《欧盟人工智能法案》对类似违规行为最高可处以1500万欧元（约合230亿韩元）的罚款相比，韩国的行政罚款对违规行为的经济威慑力较弱。

but 也很难因此就认为3000万韩元毫无意义。

重新审视施行令第29条第1款第4项。曾因违反第43条第1款第3项被处以行政罚款的人工智能经营者，此后无论销售额或使用者人数如何，均会产生指定韩国境内代理人的义务。一旦因未履行中止命令或纠正命令而被处以一次行政罚款，指定韩国境内代理人的义务便不会被解除。比起3000万韩元金额本身，我们更应关注随着行政处罚记录的累积导致法律义务不断扩大的机制。

在未指定韩国境内代理人的状态下，即使科学技术信息通信部想要实施事实调查（第40条），也没有可以触及作为调查对象的境外经营者的物理窗口。指定韩国境内代理人的义务，其核心不在于行政罚款的数额大小，而在于这是韩国政府为确保能够与境外经营者取得联系而设置的最低限度通道。

本法将适用整改过渡期。《人工智能基本法》的行政罚款征收将至少宽限一年，推迟至2027年1月以后。除非发生严重损害人权或造成人员伤亡的事故，否则

行政指导和纠正建议将予以优先。对于境外经营者而言，这一宽限期正是指定韩国境内代理人并整備申报体系的时间。

8. 实务判断流程

在此梳理境外人工智能经营者判断自身义务的步骤。

在韩国境内是否设有住所或营业场所？若有，则不属于第36条的适用对象。无需另行指定韩国境内代理人。然而，如果人工智能服务的合同当事人是境外法人，而韩国法人仅起支持作用，则该境外法人仍可能是适用对象。

若是适用对象，是否满足施行令第29条四项标准中的至少一项？上一年度总销售额达到1万亿韩元以上；人工智能服务部门销售额达到100亿韩元以上；紧邻的前3个月内韩国境内日均使用者达到100万人以上；曾因未履行中止命令或纠正命令而被处以行政罚款。只要满足其中任何一项，就必须指定韩国境内代理人。

若符合标准，应当以书面形式指定韩国境内代理人，并将该指定事实向科学技术信息通信部长官申报。韩国境内代理人应当是在韩国境内拥有住所或营业场所的人。律师事务所、咨询公司、韩国境内关联公司等均可成为代理人。

指定代理人后，义务仍将持续。韩国境内代理人将代理提交高性能人工智能安全性履行结果、申请确认高影响人工智能、以及支持履行高影响人工智能确保安全性措施等事项。如果韩国境内代理人在这些范围内违反法律，其责任将归属于境外总部。

让我们回到江南会议室的那位副总裁身上。如果有韩国分公司，可能确实不属于第36条本身的适用对象。然而，如果人工智能服务条款的合同当事人是美国总部，那么即便设有韩国分公司，总部仍可能被认定为“在韩国境内没有住所或营业场所”的经营者。法律条文是简短的，而解释的空间是广阔的。在欧盟，如果不指定授权代表，可能会面临数百亿韩元的罚款，而在韩国则是3000万韩元。虽然金额差距悬殊，但并不意味着这项义务无足轻重。韩国境内代理人是韩国政府

联络境外经营者的通道，如果该通道受阻，无论是事实调查还是纠正命令都将无法送达。该条款的本质不在于行政罚款的数额大小，而在于企业是否具备了在韩国市场继续开展业务所必需的最低限度基础设施。

第三部 所有经营者的基本义务：确保透明性

第7章. 事先告知义务：何时、何物、如何告知

2026年1月的某一天，在首尔江南的一家金融科技公司的会议室里。法务团队负责人在屏幕上展示了本公司贷款比较应用程序的屏幕截图。该应用程序分析使用者的信用信息，按利率由低到高的顺序推荐贷款商品。推荐引擎是一个机器学习模型。法务团队负责人问道：“根据韩国《人工智能基本法》第31条第1款，必须事先告知使用者正在使用人工智能（AI）。但是，‘事先’到底是什么时候？是第一次安装应用程序的时候吗？是注册会员的时候吗？还是在显示推荐结果之前？”开发团队负责人接着问道：“告知的方法是弹出窗口、条款，还是屏幕底部的附属细则？而且，只要写上‘使用AI’就可以了，还是必须写明是什么种类的人工智能，以及用什么数据进行的训练？”法务团队负责人阅读了法律条文，但没有找到答案。这不仅仅是这家公司的问题。在人工智能基本法支援台接到的在线咨询中，有56.4%是关于透明度义务的。

1. 法律的要求

韩国《人工智能基本法》第31条第1款只有一句话：“人工智能经营者在提供使用高影响人工智能或生成式人工智能的产品或服务时，必须事先告知使用者该产品或服务是基于该人工智能运行的。”

分解这句话，可以得出四个要素。

义务的主体是人工智能经营者。开发经营者和使用经营者均包含在内。不仅是亲自创建人工智能模型并提供服务的公司，引入外部人工智能并以自己的名义提供产品或服务的公司也属于此列。

作为对象的人工智能有两种。即高影响人工智能和生成式人工智能。一方面是在能源供应、医疗保健、招聘、贷款审查、交通管理、公共服务、学生评价等法律列举的十个领域和总统令规定的领域中使用的高影响人工智能。另一方面是生成文本、图像、语音、视频的生成式人工智能。对于使用不属于这两个类别的人

人工智能的经营者，不适用该条款的事先告知义务。对客户咨询进行分类的机器学习模型或物流优化算法不属于高影响领域，也不生成任何内容，因此被排除在第31条第1款的事先告知对象之外。

告知内容很明确。“产品或服务基于该人工智能运行的事实”。即使用人工智能这一事实本身。是什么模型，用什么数据进行的训练，准确率是多少，这些都不属于该条款要求的范围。

时间点是“事先”。必须在使用者使用产品或服务之前告知。

如果违反，将根据第43条第1款第1项处以3000万韩元以下的行政罚款。在韩国《人工智能基本法》中，涉及行政罚款条款的义务只有三个：未履行事先告知、未指定韩国境内代理人、未履行中止及纠正命令。事先告知是其中之一。3000万韩元这个金额对大企业是否具有威慑力尚有争议，但从法律实施的第一年起就被纳入行政罚款征收对象，这是一个明确的事实。不过，将适用至少一年的辅导期，除非发生人员伤亡或大规模侵犯人权等重大事件，否则行政指导和纠正建议将优先。

2. “事先”是指什么时候

法律第31条第1款仅写了“事先”。施行令第23条虽然列举了告知方法，但并未定义“事先”的具体时间点。这种空白似乎是有意为之。因为使用人工智能的产品和服务的形式太多样化，如果固定一个时间点，就会出现不符合现实的情况。

在实际业务中，“事先”根据服务类型可分为三种情况。

如果是需要合同的服务，那么在签订合同之前是一个自然的时间点。就贷款比较应用程序而言，在同意使用条款之前，必须包含“本服务基于人工智能推荐商品”的告知。如果在签订合同后才迟迟告知，使用者就失去了选择基于人工智能的服务的机会，这不符合“事先”的宗旨。

如果是像网站或应用程序那样无需单独合同即可访问的服务，则以首次使用时间点为基准。只要使用者在首次打开服务屏幕时能确认基于人工智能运行的事实即可。在使用服务很久之后才在设置菜单深处发现，这不是“事先”。

如果是线下产品，则是在购买前或首次使用前。搭载人工智能的家用电器的包装或门店展示中必须有告知。

《确保透明度指南》支持这种解释。虽然指南没有法律约束力，但它是科学技术信息通信部将第31条的宗旨具体化后的业务指南，因此是经营者在设计合规性时首先应参考的文件。指南提出了必须在“使用者决定使用产品或服务之前”进行告知的原则。判断的基准点是使用者的“决定”。必须保障使用者在明知是基于人工智能的服务的情况下仍选择使用的机会。

3. 四种告知手段

施行令第23条第1款列举了四种告知方法。

有直接记载在产品上，或记载在合同、使用说明书、使用条款等中的方法。在供应B2B软件时，在合同中写明“本软件基于人工智能运行”，就属于这种类型。在人工智能音箱的包装盒上印上“基于人工智能的语音识别产品”也是一样的。优点是能留下证据，并且在签订合同的过程中自然地完成了告知。缺点是，如果被埋在条款的长篇大论中，使用者实际上可能无法认知。

有显示在使用者屏幕或终端设备上的方法。弹出窗口、横幅、应用程序内通知、网页顶部告知文均属此类。例如，在首次运行贷款比较应用程序时，弹出一个写着“此应用程序提供基于人工智能的推荐服务”的窗口，并让用户按下确认按钮。从直接进入使用者视线的角度来看，这是一种确切的方法。但是，它的局限性在于，如果积累了弹出窗口疲劳（popup fatigue），使用者就会养成不读就关闭的习惯。

有在提供产品或服务的场所，以易于认知的方法进行公布的方法。括号内附带了“包括与该场所合理相关范围内的场所”的但书。例如，在无人便利店入口处张贴“本店运营基于人工智能的顾客动线分析”的告示。使用人工智能面试系统的企业在招聘公告页面上公布“面试过程中将包含人工智能分析”，也属于这种类型。

最后是有科学技术信息通信部部长认可的方法。这是一个开放条款。因为随着技术的发展，告知手段也可能发生变化，所以这是一个包容难以归入前三种类别的新方法的机制。

施行令将选择这四种方法中的哪一种交给了经营者。可以只用一种，也可以组合使用多种。但是，无论选择哪种方法，都必须让使用者能够实际认知。把内容用8号字体写在条款第47页末尾，即使在形式上属于“记载”，也无法满足指南所指的“事先告知”的宗旨。

4. 应该告知到什么程度

法律第31条第1款要求的告知内容是“基于该人工智能运行的事实”。即使用人工智能这一事实本身。该条款不会要求更多内容。

具体来说，以下信息不是第31条第1款事先告知的对象：使用什么模型（是GPT-4还是Gemini）、用什么数据进行的训练（是公开数据还是自行收集的数据）、模型的准确率或错误率是多少、算法的运行原理是什么。

“使用AI”是告知的底线。

但是，在底线之上还存在着其他义务。对于运营高影响人工智能的经营者，第34条第1款第2项规定了单独的解释义务。必须“在技术可能范围内”，制定并实施有关人工智能得出的最终结果、用于得出最终结果的主要标准、用于人工智能开发及使用的训练数据的概要等的解释方案。施行令第27条第1款规定，该解释方案的主要内容应公布在经营者的互联网站点等处。属于商业秘密的事项可以除

外。

综上所述：第31条第1款的事先告知是告知“使用AI”的事实。第34条第1款第2项的解释义务是对“人工智能如何做出决定”的解释。前者适用于所有高影响人工智能和生成式人工智能，而后者仅适用于高影响人工智能。前者在利用前履行，而后者在服务运营期间也必须继续履行。

作为基本原则条款的第3条第2款则更进了一层。“受影响者应能在技术和合理可能的范围内，获得有关得出人工智能最终结果所利用的主要标准及原理等清晰且有意义的解释。”这是受影响者必须能够获得解释的基本原则。虽然这不是附带单独请求程序或制裁的权利，但该条款阐明了一个方向，即当高影响人工智能做出对自己不利的决定时，能够质问为什么会做出这样的决定。

事先告知、解释义务、提供解释的基本原则。这三个层级的透明度机制叠加在一起，而事先告知是其中的第一层。因为如果不知道“使用AI”的事实，就既不能要求解释，也无法获得解释。

《确保透明度指南》建议经营者在编写事先告知文句时，尽可能做到清晰易懂。与其使用“基于人工智能”这种表达，不如像“本服务由人工智能（AI）提供推荐结果”那样，将服务的具体功能一并写明，这样更为妥当。指南中还有避免使用专业术语，并考虑使用者年龄和数字熟练程度的指导方针。施行令第23条第3款在涉及深度伪造（Deepfake）的告知中规定，应“考虑主要使用者的年龄、身体、社会条件等”来设计告知方式，这一原则在一般的事先告知中也值得参考。

5. 生成物标识与深度伪造标识

在同一条文中，还有两个必须与第31条第1款的事先告知区分开来的义务。

第31条第2款是生成物标识义务。提供生成式人工智能或利用它提供产品或服务的经营者，必须标明其结果是由生成式人工智能生成的事实。如果说事先告知是对服务本身的告知，那么生成物标识则是对每一个生成结果的标识。在人工智

能图像生成服务生成的照片上贴上“AI生成”的标签，这就是生成物标识。

施行令第23条第2款将标识方法分为两种。即人可认知的方法（可见标签、文本、语音引导）和机器可读的方法（元数据、数字水印）。在使用机器可读方式时，必须通过引导文本或语音等提供至少一次该结果由生成式人工智能生成的事实。仅仅使用只有机器才能读取的水印是不够的。

第31条第3款在程度上更为严格。在利用人工智能系统生成难以与现实区分的虚拟声音、图像、视频的情况下，必须以使用者“能明确认知的方式”进行告知或标识。这是针对深度伪造的条款。但属于艺术或创意表达物的情况，可以采取不妨碍展示或欣赏的方式进行告知。这意味着如果在人工智能电影的某一场景全屏打上“AI生成”的水印会破坏作品，因此可以通过在片尾字幕或作品说明中标示来代替。

施行令第23条第3款在深度伪造标识上附加了两个标准。一是用能通过视觉、听觉或软件轻易确认内容的方法进行标识。二是应考虑主要使用者的年龄和身体、社会条件等。这其中蕴含的意义是，如果是老年人经常使用的服务，就应该放大字体或用语音进行引导。

6. 豁免告知的情况

施行令第23条第4款规定了三种例外。

考虑到产品或服务的名称、使用者屏幕、产品外观、结果上标示的文本等，使用高影响人工智能或生成式人工智能的事实显而易见的情况。像“ChatGPT”、“DALL-E”、“AI翻译器”这样名称本身就展现出是人工智能的服务，就无需另行告知。因为使用者在登录时已经知道这是基于人工智能的服务，所以告知的目的已经达到。

仅供人工智能经营者内部业务使用的情况。员工在公司内部将人工智能用于内部分析或决策支持，但只要其结果不直接提供给外部使用者，就没有事先告知义

务。但是，对这个例外的解读应当狭义。如果是以人工智能分析结果为基础对使用者做出决定（例如用人工智能筛选简历后向申请人发送合格或不合格通知），该人工智能就不属于“仅供内部业务使用”。因为结果会对使用者产生影响。

此外，综合考虑产品或服务的类型和特点、结果的内容、使用形式、技术水平等因素，科学技术信息通信部部长通过公告规定需要适用例外的情况。截至2026年7月，该公告尚未发布。

在阅读例外规定时需要注意一点。该例外并不局限于事先告知（第31条第1款），还可以包含生成物标识（第2款）和深度伪造标识（第3款），适用于其全部或部分。但是，即使属于例外情况，也不意味着能够摆脱整个透明度义务。高影响人工智能经营者的解释义务（第34条）或风险管理义务仍将单独保留。

7. 欧盟：所有AI系统的透明度

欧盟《人工智能法案》（EU AI Act）第50条在比韩国法律更广的范围内要求透明度。

第50条第1款。如果人工智能系统被设计为直接与自然人互动，provider（开发供应商）必须告知该自然人，他正在与人工智能系统互动。“在情况和使用背景下显而易见的情况”属于例外。也就是说，聊天机器人在与客户对话时，必须声明“我是AI”。这与韩国法律是同样的原理，但适用范围不同。韩国仅将高影响人工智能和生成式人工智能作为告知对象。而欧盟则是针对所有与自然人互动的人工智能系统。无论其风险等级如何。

第50条第2款。生成合成音频、图像、视频、文本内容的人工智能系统的provider，必须在结果上以机器可读格式（machine-readable format）标识人工智能生成的事实。韩国法律承认机器可读方式，但仅靠它是不够的，必须同时提供至少一次人可认知的引导。欧盟将机器可读标识作为provider的基本义务，而将人可认知的公开（disclosure）交给了deployer（部署使用者）。

第50条第3款。情感识别系统或生物分类系统的deployer，必须向暴露在该系统下的自然人告知该系统的运行事实。韩国法律中没有对情感识别人工智能的单独告知义务。

第50条第4款。生成或操作属于深度伪造的图像、音频、视频内容的人工智能系统的deployer，必须公开该内容是人为生成或操作的事实。如果是使用人工智能生成或修改旨在通报有关公共利益事项而发布的文本，也有公开义务。但是，如果人工智能生成的内容由人进行审查并承担编辑责任（例如新闻社记者修改人工智能初稿后发表文章），则免除文本公开义务。

韩国法律中没有这种“编辑责任”豁免。原则上，必须对生成式人工智能生成的结果进行标识。

韩国看重影响。欧盟看重互动。韩国法律根据人工智能的类型（高影响、生成式）划分了透明度义务。欧盟法律则根据人工智能系统与人接触的方式划分了义务。结果是，欧盟的法网更广。在韩国不是事先告知对象的低风险推荐算法，如果在欧盟与使用者直接互动，就会成为告知对象。

罚款差距依然存在。韩国违反事先告知的行政罚款最高为3000万韩元。根据欧盟《人工智能法案》第99条，对违反第50条的罚款最高可达1500万欧元或全球年营业额的3%，以两者中较高者为准。这是750倍以上的差距。

欧盟《人工智能法案》的透明度义务（第50条）将从2026年8月2日起适用。因为韩国《人工智能基本法》的透明度义务（第31条）已于2026年1月22日施行，所以欧盟方面晚了约6个月。这两部法律几乎在同一时期生效。欧盟虽然决定通过数字法案综合修订将高风险人工智能义务推迟至2027年12月，但保留了聊天机器人公开和深度伪造标识等透明度条款的时间表。这是把透明度视为不可推迟的义务。

8. 中国：备案、标识和控制

中国的人工智能透明度义务在设计上与韩国或欧盟不同。透明度条款不是包含在一部综合性法律中，而是散落在各领域的规定里。

《互联网信息服务算法推荐管理规定》（2022年3月施行）赋予了算法推荐服务提供者告知使用者的义务。必须告知使用者正在提供算法推荐服务的事实，并保障使用者可以关闭个性化推荐的选择权。到这里为止，与韩国法律的事先告知相似。不同之处在于备案。具有舆论属性或社会动员能力的服务提供者，必须将其算法的类型、应用领域、自评估结果向国家互联网信息办公室（CAC）备案。自开始提供服务之日起10个工作日内进行备案，若备案内容发生变更，则在10个工作日内再次办理变更备案。在韩国，没有向政府登记人工智能的义务。

《生成式人工智能服务管理暂行办法》（2023年8月施行）更进了一步。若要向大众提供具有“舆论属性或者社会动员能力”的生成式人工智能服务，在此之前必须申报安全评估并完成算法备案。服务提供者必须向使用者告知生成内容是由人工智能生成的事实，并且必须遵守“社会主义核心价值观”。韩国法律中没有对内容本身进行控制的规定。

2025年9月1日施行的《人工智能生成合成内容标识办法》将标识义务进行了精密的规范化。服务提供者必须在人工智能生成的内容中添加机器读取的隐式标识（元数据、数字水印），而对于像深度伪造那样可能引起混淆的类型内容，还必须同时配备使用者肉眼可见的显式标识（可见标签、文本语句）。强制性国家标准 GB 45438-2025 统一了标识的格式和规范。韩国法律将标识方法交由经营者选择，并未规定具体规范。而在中国，国家甚至规定了标识规范。

备案现状显示了该制度的规模。截至2024年8月，已有190余个大型语言模型完成了备案，到2025年11月，已有611个生成式人工智能服务进行了备案。据悉，全部算法备案数量已超过5000件。

相关规定至今仍在增加。2026年4月由国家互联网信息办公室（CAC）等五个部门共同公布、并于同年7月15日起施行的《拟人化人工智能交互服务管理规定

》，针对像人一样对话并交流情感的人工智能服务，设立了必须向使用者明确表明对方是人工智能而非人类的亮明身份义务。这比在内容上添加标识更进一步，其宗旨在于防止人工智能伪装成人类来动摇使用者的情感和判断。

综合这三个国家来看，透明度的含义各不相同。韩国是告知使用者。欧盟是告知使用者，并在结果上进行标识。中国是告知使用者，在结果上进行标识，并向政府备案。韩国的透明度面向使用者。中国的透明度则同时面向使用者和国家。

在执法力度上也存在差异。韩国最高3000万韩元的行政罚款，在辅导期间实际上仅止步于行政指导。据传，中国在2025年互联网主管部门的人工智能技术滥用整治行动中，删除了数十万条违法内容，并处置了大量人工智能账号和服务。随后会采取服务中断这种事实上的强力制裁。

9. 业务设计问题

在阅读了法律条文、施行令和指南之后，摆在经营者面前的是设计问题。

告知文句该怎么写。“本服务基于人工智能运行”这一句话是最低限度。但是，这句话是否能为使用者提供有意义的信息，则是另一个问题。如果是贷款商品推荐应用程序，写上“此应用程序由人工智能分析您的信用信息并推荐贷款商品”会有助于使用者的理解。指南也建议将告知与服务的具体功能结合起来。

该布置在何处。在法律上，在条款里加一句话也是可以的。在实际业务中，将其布置在服务屏幕上使用者视线所及的地方，能预防争议。因为在争议中，“使用者实际认知了”比“已经告知了”这一事实更有可能成为争论焦点。是在注册会员阶段的同意屏幕、服务主屏幕的顶部横幅，还是结果提供屏幕的附加说明中更为合适，则因服务的性质而异。

告知一次就结束了吗。法律只规定了要“事先”告知，并未规定频率。法律中没有一旦告知后必须重复的义务。但是，如果服务更新导致人工智能的作用发生变化，或者以前未使用人工智能的功能新加入了人工智能，就必须重新告知。需

要发布“此功能已更改为基于人工智能”的更新告知。

如果不确定是否属于高影响人工智能，该怎么办。向科学技术信息通信部申请确认的话，会在30日内收到结果回复。可延长一次，为期30日。如果对确认结果有异议，可以在10日内要求重新确认，重新确认将在经过专家委员会咨询后的30日内出具。在确认程序进行期间，如果已经正在提供服务，最好还是进行事先告知以策安全。即使结果显示不是高影响人工智能，告知了也不会有什么损失。

对所有这些问题，并没有一个唯一的标准答案。第31条和施行令第23条只是设定了框架，而框架内的具体执行则是经营者的任务。截至2026年3月31日，人工智能基本法支援台收到的552件咨询，正是经营者们在这个框架内摸索道路的足迹。

前面会议室里的法务团队负责人曾问过：“‘事先’到底是什么时候。”法律并没有规定日期。取而代之的是，它给出了一个原则：不应剥夺使用者决定是否使用基于人工智能的服务的机会。将该原则翻译成符合自身服务的内容，仍是留给经营者的课题。

第8章。生成物标识义务：水印、标签以及深度伪造的特例

2025年10月，某房地产中介平台上发布了12张首尔江南新建商住楼的房源照片。宽敞的客厅、落地窗外的汉江夜景、岛台式厨房。照片十分清晰，光线也很自然。咨询电话纷至沓来。然而，那栋商住楼当时还处于骨架工程施工阶段。这12张照片其实全部是人工智能生成的图像。售楼公司在任何地方都没有标注使用人工智能的事实。而于2026年1月22日施行的《人工智能基本法》，在当时还尚未出台。

在该法实施后，情况就不同了。该法第31条要求人工智能经营者在其生成式人工智能的输出结果上进行标识。制作出这些照片的图像生成服务经营者首先承担这一义务，而利用人工智能制作广告并使用的售楼公司是否属于该法的受规制主体，则要视具体情况而定。施行令第23条具体规定了标识的方法和例外情况。第31条将事先告知、生成物标识、深度伪造告知这三重义务凝聚在了一个法条之中。

1. 三重透明度

解析第31条的结构。

第1款是事先告知。在提供利用高影响人工智能或生成式人工智能的产品或服务时，必须预先告知使用者该产品或服务是基于相应人工智能运行的事实。这是关于“本服务使用人工智能”的告知。这是针对服务本身的标识，而不是针对每一个输出结果的标识。

第2款是生成物标识。在提供生成式人工智能或利用其开发的产品或服务时，必须标识该输出结果是由生成式人工智能生成的事实。如果说第1款是挂在服务入口处的招牌，那么第2款就是贴在服务输出的每一个产出物上的标签。ChatGPT撰写的报告、Midjourney绘制的图像、Suno AI创作的音乐，都必须分别标注“由人工智能生成”。

第3款是深度伪造特例。利用人工智能系统提供与真实情况难以区分的虚拟声音、图像或视频等输出结果时，必须以使用者能够清晰识别的方式进行告知或标识。这是比第2款更高一级的要求。第2款是“进行标识”，而第3款则是“以使用者能够清晰识别的方式进行标识”。这里多了一个修饰词。

这三重义务的适用范围各不相同。第1款的事先告知同时涵盖了高影响人工智能和生成式人工智能。无论是用于招聘筛选的人工智能系统（高影响），还是图像生成服务（生成式），都必须预先告知使用者使用了人工智能的事实。第2款的生成物标识仅适用于生成式人工智能。这仅限于人工智能并非进行判断或推荐，而是生产内容的情况。第3款的深度伪造特例甚至涵盖了非生成式的人工智能系统。由于该条款表述为“利用人工智能系统”，因此即使是非生成式人工智能，只要制作出与真实情况难以区分的声音或视频，就会成为规制对象。

2. 可见标识与机器可读标识

经营者具体如何进行标识，取决于施行令第23条的规定。

第1款列出了四种事先告知的方法。在产品上直接记载，或写在合同、使用说明书、使用条款中的方法；在使用者屏幕或终端上显示的方法；在提供服务的场所中张贴于易于识别之处的方法；以及科学技术信息通信部长官认定的其他方法。如果是人工智能聊天机器人服务，在对话界面上方加入“本服务由人工智能运行”的提示，便是可以直接想到的履行方式。

第2款将生成物标识的方法分为两种。

一种是人可以识别的方法。即可见标识。标注“人工智能生成”的字样、在图像角落插入的水印图标、在视频底部放置的字幕，都属于此列。

另一种是机器可以读取的方法。即不可见标识。例如在文件的元数据中植入“此内容由人工智能生成”的信息，或者在图像中插入肉眼不可见的数字水印。要读取元数据或水印，需要单独的软件。人无法用肉眼进行确认。

这体现了核心的设计抉择。施行令允许在两者中选择其一。其表述为“可以通过符合下列各号之一的方法进行”。既可以仅进行可见标识，也可以仅进行机器可读标识。但是，如果仅选择机器可读方法，则会附加一个条件：“必须通过提示语或语音等方式，提供一次以上该输出结果是由生成式人工智能生成的事实。”仅在元数据中隐藏信息是不够的。必须至少以人类可读的形式，向使用者告知一次由人工智能生成的事实。

理顺这一结构，情况如下。经营者可以仅通过可见标识来完成义务，或者也可以选择机器可读标识的基础上增加至少一次提示语的方式。仅单独进行机器可读标识是不允许的。

3. 深度伪造要求的放宽并不等于免除标识

第31条第3款的深度伪造特例确立了比第2款更高的标准。

“使用者能够清晰识别的方式”这一法条表述是核心所在。施行令第23条第3款将这一要求具体细分为两个方面。即使用者必须能够通过视觉、听觉或利用软件轻松确认内容；同时，还应当考虑主要使用者的年龄以及身体、社会条件。

所谓“清晰识别”，很难仅通过机器可读标识来满足。即使在元数据中植入了信息，普通使用者也不会去打开元数据查看器。即使植入了肉眼不可见的水印，只要不运行检测软件，谁也无法知晓。深度伪造的风险在于其传播速度。一旦扩散，便难以挽回。与事后用机器验证相比，事前让人眼直接可见，才是本条款的宗旨。

法律做出了一项放宽规定。如果深度伪造的输出结果属于艺术性或创造性的表达物，或者构成了其中的一部分，则可以采取不损害展览或欣赏的方式进行告知或标识。利用人工智能制作的微电影或音乐视频便属于此类情况。如果在电影正片画面上一直浮现“人工智能生成”字样，就会干扰观影。这意味着，在此类情况下，可以通过在片尾字幕中注明，或者在发布页面上予以告知来代替。然而，这并不意味着免除标识本身。

4. 适用例外的三种情形

施行令第23条第4款规定了三种例外情形。这些例外涵盖了事先告知、生成物标识、深度伪造告知这全部三项义务，在符合条件时可以不适用相关规定的全部或部分内容。

综合产品或服务的名称、使用者界面、产品外包装、输出结果上标注的字样等来看，如果使用人工智能的事实显而易见，则无需进行额外的标识。从名为“AI绘画生成器”的应用程序中导出的图像，无需再次注明“AI生成”。服务名称本身已经起到了告知的作用。这一例外是符合常理的防重复机制。

在仅用于经营者内部业务用途的情况下，也可以不适用全部或部分告知与标识义务。如果营销团队为了创意会议制作了人工智能图像并将其放入公司内部的幻灯片中，则无需进行标识。但该图像一旦被用于对外广告，该例外就会立即失效。从内部走向外部的分界线，即为义务的起点。

第三种是科学技术信息通信部长官通过公告规定的例外情况。可以综合考虑产品或服务的类型、特性、输出结果的内容、使用形式、技术水平等，设置追加的适用例外。该条款属于尚未出台具体公告的开放性授权。

5. 如果违反会怎样

根据该法第43条第1款第1项，违反第31条第1款即事先告知义务者，将被处以3000万韩元以下的行政罚款。

此处的处罚机制颇具特点。行政罚款条款仅明确列出了违反第31条第1款的情形。第43条中并没有关于违反第2款（生成物标识）或第3款（深度伪造告知与标识）的行政罚款条款。相反，第40条第1款规定，科学技术信息通信部长官在发现或接到违反第31条第2款或第3款事实的举报时，可以要求其提交资料或进行事实调查。第40条第3款规定，如果调查结果认定存在违法事实，可以下达中止命令或纠正命令。随后，根据第43条第1款第3项，不履行该中止命令或纠正命令的人，将被处以3000万韩元以下的行政罚款。

从结构上来看，未在生成物上进行标识的行为本身并不会直接导致行政罚款。只有在科学技术信息通信部进行调查并下达纠正命令后，若仍不遵守该纠正命令，才会处以行政罚款。未履行事先告知义务会直接被处以行政罚款；而未履行生成物标识义务则需要经过“不履行纠正命令”这一中间步骤。整改过渡期（至少1年，宽限至2027年1月之后）同样适用于事先告知的行政罚款。然而，对于恶意利用深度伪造犯罪或造成人身伤亡等重大案件，则可立即展开调查。

6. 欧盟：强制要求机器可读标识

《欧盟人工智能法案》第50条与韩国法律第31条处于同等定位，但切入路径有所不同。其正式名称为“针对特定人工智能系统提供者和部署者的透明度义务（Transparency obligations for providers and deployers of certain AI systems）”。该条由7个款项组成，比韩国法律的三重结构更加细密。

第50条第1款是交互告知。人工智能系统的提供者必须确保使用者知晓其正在与人工智能系统进行交互。但“从具备合理信息储备的自然人角度来看显而易见的情形”除外。这与韩国法律第31条第1款的事先告知相对应，但欧盟直接在法律条款中写入了“显而易见情形下的例外”，而韩国则在施行令中对该例外做出了规定。

第50条第2款是合成内容标识。该条款要求人工智能系统提供者（provider）以机器可读格式（machine-readable format）标识合成的音频、图像、视频或文本内容的输出结果已被人工生成或篡改。技术解决方案必须“有效、可互操作、强健且可靠（effective, interoperable, robust and reliable）”。序言（Recital）第133项列举了允许使用的技术，包括水印、元数据识别、证明来源和真实性的密码学方法、日志记录方法、指纹技术等。C2PA恰好符合该列表中所提到的密码学方法与元数据识别技术。这与韩国法律允许经营者在可见标识和机器可读标识之间进行选择形成了对比。欧盟将机器可读标识设为了基本义务。仅履行辅助编辑功能的人工智能系统，或不实质性改变输入数据的人工智能系统，则可免除该义务。

第50条第4款是针对深度伪造和人工智能生成文本的独立义务，将两者捆绑在了一起。当利用人工智能系统生成或篡改图像、音频、视频，使真实存在的人物、事物、地点或事件看起来像真实发生的一样时，必须公开该内容已被人工生成或篡改的事实。该义务被赋予部署者（deployer）。出于执法目的使用的情况、属于讽刺或滑稽模仿等言论自由范畴的情况、以及属于艺术创作的情况均属例外。即使是艺术性表达物，也不能免除标识本身，而必须以“不损害作品展览或欣赏的适当方式（an appropriate manner that does not hamper the display or enjoyment of the work）”进行标识。这与韩国法律第31条第3款关于艺术性表达物的放宽规定有着几乎相同的结构。不同之处在于，欧盟法案明确将讽刺和滑稽模仿纳入了例外。韩国法律中则没有这类明文免除条款。该款后半部分还涉及了出于公益信息目的的人工智能生成文本。若经过人工编辑审查并由其承担编辑责任，则可免除公开义务。该规定考虑到了新闻媒体将人工智能作为辅助工具，同时由记者行使最终编辑权的情况。

第50条第7款将促进制定行为准则的权力授予了欧洲人工智能办公室（AI Office）。其旨在制定欧洲层面的用于检测和标识人工智能生成内容的行为准则，并允许欧盟委员会通过实施法案（implementing acts）。

欧盟与韩国在结构上的差异如下。

韩国侧重于评估影响，欧盟则侧重于评估风险。这一差异在标识义务上同样有所体现。

韩国第31条第2款以“应当进行标识”结尾。具体的方法由经营者在施行令规定的可见标识或机器可读标识中自行选择。而欧盟第50条第2款则将机器可读格式列为强制义务。韩国赋予了经营者选择权，欧盟则直接指定了技术实现方式。

韩国的深度伪造条款（第31条第3款）要求“使用者能够清晰识别的方式”，这在实际上强制实施了可见标识。欧盟的深度伪造条款（第50条第4款）则要求予以“公开（disclose）”。双方都认识到，对于深度伪造而言，仅靠机器可读

标识是远远不够的。

《欧盟人工智能法案》的透明度义务已于2026年8月2日起开始适用。尽管由于数字综合修正案的通过，高风险人工智能义务被推迟至2027年12月，但透明度义务（第50条）仍按原定计划施行。综合修正案并未对第50条本身做出修改。对于在2026年8月2日之前已经投入市场的系统，给予了4个月的宽限期。

行政罚款的差距在该条款中也得到了印证。欧盟对于违反第50条的行为，最高可处以1500万欧元或全球年营业额3%的罚款。而韩国在未履行纠正命令时，最高仅处以3000万韩元罚款。

7. 中国：确立双重标识体系

2025年9月1日，中国施行了《人工智能生成合成内容标识办法》。该办法由国家互联网信息办公室（CAC）、工业和信息化部、公安部、国家广播电视总局四个部门联合制定。

韩国是在可见标识和机器可读标识中选择其一。中国则是将双重标识层层叠加。

即显式标识与隐式标识的双重体系。显式标识是使用者肉眼可见的标签。例如在图像中添加“AI生成”字样，或在视频播放过程中在屏幕上显示人工智能生成标识。隐式标识则是植入在文件元数据中的用于机器读取的信息。

这双重标识的适用范围各不相同。隐式标识，即在文件元数据中植入人工智能生成信息，是贯穿于生成合成内容整体的义务。显式标识则被要求用于可能导致公众混淆或误认的输出内容的服务中。而在内容本身植入数字水印则尚未强制，仍停留在鼓励层面。此外，禁止任何人恶意删除、篡改、伪造或隐匿这些标识。原文“不得恶意删除、篡改、伪造、隐匿”明确了这一禁令。

此外，国家强制性技术标准 GB 45438-2025 也于同日施行。该标准规定了标识方法的具体技术细节。它不是推荐性标准，而是强制性标准。独立于 C2PA 或 ISO 等国际标准，在中国境内传播的人工智能生成内容的标识必须符合该国家标

准。

将这三个国家并列对比，监管强度的差异便一目了然。

中国将元数据隐式标识设为针对所有内容的强制性义务，并在存在混淆隐患的服务中增加了显式标识要求，甚至制定了国家强制性标准。欧盟强制要求进行机器可读标识，但仅针对深度伪造追加了公开义务。韩国则允许经营者在可见标识和机器可读标识中任选其一，仅针对深度伪造要求具备明晰的可识别性，从而在事实上引导其采用可见标识。

韩国之所以提供选择权，是考量了技术发展速度和产业负担后做出的抉择。在机器可读标识技术尚未成熟的阶段，如果强制推行特定的技术路径，可能会将过度成本转嫁给经营者。相反，中国展现出了通过国家标准来管控具体技术路径的意图。欧盟则介于两者之间，确立了“符合最前沿技术的机器可读格式”这一原则，但将具体技术方案交由行为准则（Code of Practice）来解决。

8. 水印技术能否兑现法律的承诺

法律要求进行标识，而技术是落实该标识的手段。截至2026年7月，水印技术是否已经达到了法律所要求的水平？

C2PA（Coalition for Content Provenance and Authenticity）是由Adobe、Arm、BBC、Intel、Microsoft和Truepic于2021年2月联合创立的内容来源和真实性联盟。其技术规范（specification）的2.1版本于2024年9月公开，此后历经多次迭代更新，在2026年已演进至2.4版本。C2PA会在文件内部植入一个名为清单存储区（manifest store）的结构。该结构中包含内容生成工具、编辑历史、是否由人工智能生成等信息，并使用 X.509 证书进行签名。硬绑定（hard binding）通过将整个文件打包为哈希值来检测是否存在篡改；软绑定（soft binding）则采用感知哈希（perceptual hash）或指纹技术，以便在文件重新编码后依然能够追溯其来源。

Adobe已将 Content Credentials 整合至 Photoshop、Lightroom 和 Firefly 中。Microsoft也将其应用到了 Bing Image Creator 和 Designer 里。OpenAI在 DALL-E 3 和 GPT-4o 的图像输出中植入了 C2PA 元数据。相机制造商尼康（Nikon，如Z9等）、索尼（Sony，如A9III等）、徕卡（Leica，如M11-P）也搭载了支持 C2PA 的固件。谷歌宣布将于2024年在搜索结果中显示 C2PA 信息。此外，将 C2PA 技术规范推广为国际标准的工作也正在推进之中。

Google DeepMind的 SynthID 则另辟蹊径。它直接在内容本身植入肉眼不可见的水印。目前已支持图像（Imagen）、文本（Gemini的Logit采样偏差）、音频以及视频。其设计旨在确保在经过压缩或裁剪后仍能在一定程度上留存。Meta开发了用于图像的 Stable Signature 以及用于视频的Video Seal。

然而，技术的局限性是显而易见的。

C2PA元数据极为脆弱。截图操作即可导致元数据丢失。上传至社交媒体平台时，EXIF和XMP信息大都会被剥离。尽管软绑定缓解了这一问题，但在面对强力裁剪或对抗性修改时依然脆弱无力。不可见水印（如SynthID方式）则面临着另一种两难境地。如果增强水印植入力度，会降低内容品质；如果减弱植入力度，则容易被高斯噪声、JPEG重复压缩、基于扩散模型的净化（diffusion-based purification）等攻击手段所清除。

没有任何单一技术能够满足法律要求的所有条件。C2PA元数据虽拥有广泛的生态支持，但只有在整个流转路径中保持元数据不丢失才具有意义，如果主流社交媒体平台未能予以留存，该技术将形同虚设。而不可见水印仅属于概率性检测器（probabilistic detector），而非密码学证明（cryptographic proof）。

在C2PA元数据中加入不可见水印，并结合平台级注册表（registry）的混合模式，是目前公认可行的发展方向。然而，标准化工作以及平台间的互操作性（interoperability）尚未完全实现。

韩国施行令第23条第2款规定在选择机器可读方法时必须附加“提供一次以上的提示语或语音等”这一条件，这正反映出监管层认识到仅凭机器可读技术无法为用户提供充分保护。在技术能够完全兑现法律的承诺之前，人眼直接可见的标识这一模拟机制将继续发挥安全网的作用。

9. 经营者实务清单

综合法律及施行令，生成式人工智能经营者需要核对的项目如下。

是否已在服务入口处告知基于人工智能运行的事实。无论是通过使用条款还是屏幕显示，都必须确保使用者在开始使用服务之前，能够知晓有人工智能介入的事实。这对应于该法第31条第1款和施行令第23条第1款。一旦违反，将直接面临3000万韩元以下的行政罚款处罚。

是否已在人工智能生成的输出结果上进行了标识。必须在可见标签和机器可读水印中选择其一。如果仅选择了机器可读标识，是否已提供了至少一次提示语。这对应于该法第31条第2款和施行令第23条第2款。

输出结果是否为与真实情况难以区分的虚拟声音、图像或视频。若是，是否已采用使用者能够清晰识别的方式进行标识。仅靠机器可读标识很难满足这一标准。这对应于该法第31条第3款和施行令第23条第3款。

服务名称或界面上使用人工智能的事实是否已显而易见。若是，则可能免除标识义务。是否仅用于内部业务用途。若是，则可以不适用全部或部分告知与标识义务。这对应于施行令第23条第4款。

是否属于艺术性或创造性的表达物。若是，则可以采取不损害展览或欣赏的方式进行告知或标识。标识本身不能免除。这对应于该法第31条第3款的后半部分。

10. 信任不可见之物

2023年1月10日起施行的中国《互联网信息服务深度合成管理规定》是较早对合成内容强制要求进行可见标识的规定之一。韩国《人工智能基本法》中的深度

伪造条款则在三年后才得以施行。而《欧盟人工智能法案》的透明度义务比韩国晚了7个月，于2026年8月起实施。

三个国家面临着相同的问题。当人工智能生成的内容与人类制作的内容难以区分时，该边界由谁来标示、又该如何标示？各国的解决方向基本一致：由提供者进行标识，使用者则根据标识进行判断。此外，中国还禁止了擦除或修改标识的行为。

技术层面的解决方案目前仍未完善。C2PA的元数据只需截一次图就会丢失，SynthID的水印在精细的攻击下也会被剥离。法律的要求与技术可实现的水平之间尚存差距。

填补这一差距，归根结底需要平台的协同合作。如果社交媒体决定在上传过程中保留元数据，C2PA就能存活下来。反之，如果不予保留，即使投入数十亿韩元植入水印，也会在传播过程中丢失。法律强制要求经营者进行标识，但标识能否最终传达给使用者，取决于法律之外的技术生态系统。

售楼处使用人工智能生成的商住楼照片，如今也将套上标识的锁链。制作出该照片的图像生成服务经营者，必须在照片上贴上“AI生成”标签，或在文件中植入机器可读水印并显示一次提示语。对于将人工智能作为工具并发布广告图像的售楼公司而言，相比于这部法律，可能首先面临《表示广告法》等其他法规的审视。当那张照片从房地产应用程序流转到KakaoTalk，再从KakaoTalk流转到博客，进而从博客流转到新闻报道时，标签是否能够随之流转？水印是否能够完好保留？截至2026年7月，面对这些疑问，技术尚未给出确切的答案。

第九章 透明度指南实务、封闭网与开放网及例外事由

这是2026年3月的一个下午，位于首尔板桥的一家HR科技初创公司的会议室里。这家公司制作了基于AI的简历分析工具，以B2B形式交付给企业客户。该工具仅在客户公司的内部封闭网中运行。如果人事主管上传应聘者的简历，AI就会评估职务匹配度分数并推荐面试对象。应聘者并不知道AI正在阅读自己的简历。法务主管将科学技术信息通信部（以下简称“科技情通部”）的《人工智能透明度保障指南》打印出来放在了桌子上。“对内部业务用予以例外认可。”这样一句话映入眼帘。这个工具只在客户公司内部运行，看起来像是内部业务用。但是，招聘是韩国《人工智能基本法》第2条第4项列举的高影响领域。AI做出的判定决定了应聘者的合格与淘汰。虽然在封闭网内运行，但结果会触及外面的人。这适用例外吗？对于这个问题，指南里的一句话无法给出答案。

1. 事先告知与生成物标识

《人工智能基本法》向经营者赋予的透明度义务包含在第31条这一个条文中。即事先告知和生成物标识。

第31条第1款是事先告知义务。在向使用者提供使用高影响人工智能或生成式人工智能的产品或服务时，必须事先告知该事实。“该服务使用人工智能”这样的文句是典型的履行方法。其目的是让使用者认识到自己正在与AI进行交互。

第31条第2款和第3款是生成物标识义务。在生成式AI制作的结果物上必须贴上由AI生成的标识。标识既可以通过人能识别的方法进行，也可以通过机器能判读的方法进行。但是，对于难以与实际区分的虚拟声音、图像和视频，则要求采用使用者能够明确认识到的方式进行告知或标识。透明度指南建议对这种深度伪造性结果物进行肉眼可见的显性标识。因为如果没有专业工具就无法确认水印，所以判断认为仅凭不可见的水印无法对普通使用者认识到这是AI生成物提供实质性帮助。模仿、讽刺等艺术性、创意性表现物并非免除标识，而是可以通过不妨碍展示或享受的方式进行告知或标识。只是标识方式被放宽了而已。

具体解答在实务中如何履行这两项义务的，正是科学技术信息通信部的《确保人工智能透明度指南》。它于2026年1月22日法律实施的同时被发布。这是五大指南（透明度、安全性、高影响AI判断、经营者责任、AI影响评估）之一。

指南的法律性质不同于公告。它没有法律约束力。但它是事实上提出了经营者履行义务之方法的指引。当科学技术信息通信部日后判断是否履行义务时，是否遵循了指南将成为一个标准。两件公告（《确保最前沿人工智能安全性公告》、《高影响人工智能经营者责任公告》）具有法律约束力，但在透明度方面只出台了指南。透明度领域是更多地交由经营者自主决定的结构。

例外情况不是由指南，而是由施行令第23条第4款设定的。包括通过产品及服务名称、使用者界面或结果物文句能明显看出使用AI事实的情况；仅作为人工智能经营者内部业务用途使用的情况；以及由科学技术信息通信部长官通过公告规定的情况。指南处于对这些例外进行解释说明的位置。前面的两个例外在实务中引发了频繁的争议。

2. 支持台，552件咨询

AI基本法支持台于2026年1月22日开启。由韩国软件产业协会（KOSA）运营，科学技术信息通信部、NIA、TTA、KISDI和人工智能安全研究所（AISI）共同参与。

实施仅仅十天就受理了172件咨询。在实施的第10周即3月底，累积达到了552件。占累积咨询大约一半的51%是关于透明度的，确认是否属于高影响AI的咨询以19.6%紧随其后。每两个咨询中就有一个是问“应该在何时、何地、告知什么”的问题。

关于透明度的咨询如此之多是有原因的。高影响AI的分类或高性能AI的计算量标准仅适用于特定领域的经营者。如果不进入能源供应、贷款审查、生物识别等11个领域，高影响AI就与己无关。但是，透明度义务适用于所有以高影响人工智能或生成式人工智能提供产品或服务的经营者。无论是自有模型还是调用外部

API，只要服务中包含生成式AI，就必须告知使用者。除非属于例外情况。像聊天机器人、文档摘要、图像生成等包含生成式AI的服务是不分行业的。因此，大企业、中坚企业和初创企业都在纷纷提出问题。

支持台在工作日的上午9点至下午6点提供咨询服务。深入的法律审查最多需要14天。咨询内容原则上是不公开的。因为如果要询问自家服务是否属于高影响AI，就必须公开部分技术结构，所以能在保护商业机密的同时接受咨询，对经营者来说是重要的安全保障。科学技术信息通信部以实施第10周为基准发行了案例集。

3. 封闭网例外的内与外

内部业务用AI被认可享有透明度告知例外。这是施行令第23条第4款第2项规定的例外。逻辑是这样的：事先告知是为了保护使用者而设定的义务，如果使用者只有公司内部员工，且他们已经知道正在使用AI的事实，那么追加告知就没有实际利益了。

内部业务用的典型情况是这样的：公司内部会议记录摘要AI、内部文档检索辅助工具、代码审查自动化系统、营销部门使用的内部用文案生成工具。员工使用，结果也在公司内部消费。无论是封闭网还是基于云端，访问都被限制在公司内部，而且AI的产出物不会被传达给外部客户或第三方。在这种情况下，就可以适用例外。

非公开研究用模型也包含在内。如果大学实验室制作了AI模型仅用于论文实验，那么该实验室就不是人工智能经营者。根本不会产生透明度义务。但是，如果同一个实验室开始将该模型持续提供给医院或企业，那么无论是否有研究用的标签，经营者的义务都会随之而来。

到这里为止都很明确。问题在于在封闭网中运行但结果触及外部的情况。

以贷款审查辅助AI为例。该AI仅在银行的内部系统内运行。并未连接到互联网。当贷款顾问输入客户信息时，AI会计算出信用风险分数。客户无法访问该AI。在

技术上，它是封闭网。

但是，该AI被用于“对招聘、贷款审查等个人权利和义务产生重大影响的判断”这一高影响领域（第2条第4项G目）。AI计算出的分数左右着客户的贷款是否被批准。即使在封闭网内运行，结果也会直接触及客户的权利。

在这种情况下很难主张适用内部业务用例外。法律第31条第1款的事先告知想要保护的并非是能否访问AI系统。而是受AI影响之人的知情权。在法律条文上，告知的对方是接受服务的利用使用者，像贷款申请人这样受影响的人是否总是属于此列，目前还没有权威解释。但是，如果认为银行向客户提供了基于AI的贷款服务，客户就有可能属于使用者。无论银行是在封闭网还是云端运行AI，贷款申请人都处于应该知道AI正在介入其信用判断这一事实的位置。仅凭封闭网这种技术部署方式，很难被免除这项义务。

同样的逻辑也适用于招聘AI、保险金审查AI和学生评估AI。即便是仅在客户公司内部运行的封闭网AI，如果其结果决定了外部人员的合格、支付和等级，也很难受到内部业务用例外的保护。

根据AI结果有无对外影响来判断内部业务用例外，而不是根据AI的部署形式（是否为封闭网），这样比较安全。结果停留在公司内部，就是例外。结果触及外部人员的权利，就不是例外。

4. 服务名称给出答案的情况

有些情况下，服务名称本身就凸显了AI。例如“ChatGPT”、“AI翻译机”、“Naver Clova摘要”、“AI音箱”。使用这些服务的人已经知道AI正在运行。指南免除了“通过服务名称或形式能明显看出使用AI事实的情况”下的追加告知。

要适用这一免除规定，必须是“明显的”。标准在于普通使用者是否仅看服务名称就能判断出这是AI服务。

“ChatGPT”是很明显的。它是AI聊天机器人这一事实已广为人知。“AI翻译机”也很明确。名字里就包含有AI。即使没有追加告知，使用者也会认识到自己正在使用AI。

“Copilot”处于边界线上。微软的Copilot在科技界成了代表AI的名称，但对于非专业使用者来说，很难认为Copilot这个英语单词就意味着人工智能。也会有人联想到航空术语中的副驾驶。在这种情况下，比起依赖名称，加入单独告知会更安全。

还有更棘手的案例。就是AI功能被插入到现有服务中的情况。谷歌搜索结果的顶端会显示AI综合了多个来源的摘要（AI Overviews）。服务名称是“谷歌搜索”。服务名称中并没有AI字样。使用者可能很难区分传统搜索结果和AI生成的摘要。Naver搜索中默认包含AI摘要时，也会产生同样的问题。服务名称是

“Naver搜索”，但AI生成的内容却与现有搜索结果混合在一起。在这样的结构下，很难适用明显情况下的免除规定。必须在相关功能旁边单独加入由AI生成的标识。

标准可以这样来定。如果整个服务都由AI驱动且名称凸显了这一点，那么免除追加告知的可能性就很高。如果AI功能融入到了现有服务中，且结构上让使用者难以认识到，那就很难期待获得免除。AI隐藏得越深，透明度义务就越重。

5. 开放API与封闭网交付

一个AI模型可以通过两条途径提供。即通过开放API公开的情况和直接安装在客户公司内部服务器上的封闭网交付。透明度义务的适用也会有所不同。

这是开放API的结构。开发经营者（A公司）以API形式提供AI模型。使用经营者（B公司）调用该API制作自己的服务。最终使用者（消费者）使用B公司的服务。透明度义务是双层运作的。

A公司在API文档和服务条款中描述AI模型的特性、主要标准和已知局限。这种信息提供属于合同和文档的领域，如果是高影响AI，则与法律第34条的解释方案及文档化责任相衔接。法律第31条的事先告知是针对最终使用者的。B公司必须向最终使用者告知其服务中正在使用AI的事实。可以通过在聊天机器人界面上加入“这是由AI生成的回答”的提示，或者在用户协议中明示AI使用事实的方式。这是一种A公司通过合同和文档为B公司提供支持，B公司则向最终使用者进行告知的双重结构。

封闭网交付的结构不同。A公司将AI解决方案安装在B公司内部服务器上。并未连接到外部互联网。只有B公司员工才能访问。A公司在合同和技术文档中记载模型的结构和主要标准，向B公司提供必要信息。

B公司的义务取决于该AI结果所触及的范围。如果仅由B公司员工使用且结果也在公司内部消费，那么可以适用内部业务用例外。如果B公司以该AI结果为基础向外部客户提供服务，或做出了影响客户权利的判断，那么B公司就必须向客户告知AI使用事实。在封闭网中运行的事实并不能免除这项义务。

即使是同一个AI模型，如果是通过开放API提供的，由于使用者直接与AI接触，所以告知对象很明确。如果是通过封闭网交付的，在AI和最终使用者之间会有一个作为中间层的员工，因此要向谁告知什么就会变得模糊。在实务中，即使是封闭网AI，也要先确认其结果是否对外部客户产生影响，如果有影响，就必须准备与开放API同等水平的透明度告知。

还有一种变形情况。那就是存在API但访问受限的情况。这是只有内部员工或合同伙伴才能调用的非公开API。技术上虽然是API，但使用者范围却接近封闭网。在这种情况下，原则也是一样的。标准在于API结果是否会被传达给外部客户，而不是API是否公开。

6. 在B2B SaaS中由谁来告知

以SaaS（Software as a Service）形式提供AI的业务形态正在增加。A公司制作基于AI的客户分析SaaS供应给B公司，B公司则向自己的消费者提供定制服务。在消费者看来，只看到了B公司的服务。他们并不知道A公司的AI正在后台运行。

透明度告知义务在谁身上？

在《人工智能基本法》的结构中，A公司是开发经营者，B公司是使用经营者，消费者是使用者。对使用者的事先告知义务在于直接提供服务的主体。由于是B公司向消费者提供服务，所以向消费者告知AI使用事实是B公司的责任。A公司只要通过合同和技术文档，充分向B公司传达AI模型的特性和运作原理即可。

法律仅限于高影响AI责任，设有防止双重负担的机制。如果开发经营者（A公司）已经履行了风险管理、说明和使用者保护方案，且使用经营者（B公司）并未进行显著改变本来目的或用途的重大功能变更，那么就视为B公司也履行了该措施。这是施行令第27条第3款的视同规定。这种视同仅适用于法律第34条第1款第1项至第3项的高影响责任，并不适用于第31条的事先告知和生成物标识。向最终使用者告知的责任仍然留在B公司身上。施行令第27条第4款只是规定了当B公司请求时，A公司有努力协助提供资料的义务。A公司提供的技术文档和告知文句样本虽然可以成为B公司履行告知的素材，但无法代替告知义务本身。

问题在于B公司定制化A公司AI的情况。如果A公司提供了通用客户分析模型，但B公司用本公司数据进行微调并添加了行为预测功能，这就可能属于“重大功能变更”。在这种情况下，B公司有可能会超越使用经营者而被重新分类为开发经营者，并且失去了视同规定的保护，甚至需要独自履行高影响责任。关于微调到何种程度属于“重大功能变更”，截至2026年7月，科学技术信息通信部还没有给出权威解释。

实务上的解决办法是在B2B SaaS合同中明示透明度义务的归属。在合同中规定由谁向最终使用者进行告知，由谁准备技术文档，以及当AI功能发生变更时义务将如何转移。法律规定的义务归属与合同规定的业务分担是两码事。虽然法律责

任留在使用经营者B公司身上，但实际的告知文句撰写和技术文档准备可以通过合同交由A公司负责。如果忽略了这种区分，A公司和B公司都会觉得“对方会做的吧”，从而产生谁都不告知的空白。

还有一点。那就是多家客户公司同时使用A公司SaaS的情况。每家客户公司的行业不同，应用AI的领域也不同。同样是SaaS，金融公司将其用于贷款审查，流通公司将其用于客户推荐，医疗机构将其用于患者分类。在金融公司和医疗机构使用的情况下，因为属于高影响领域，所以透明度义务会被加重；但在流通公司的情况下，只要进行一般的事先告知即可。在A公司看来，只是交付了同样的产品，透明度义务的水平却会根据客户公司的用途而发生变化。A公司认识到自己的产品可能会在高影响领域中使用，并提前准备好与之相符的透明度资料，这样做会比较安全。

7. 欧盟划分范畴，中国盖章打戳

《欧盟AI法案》第50条按范畴划分了透明度义务。

必须向与AI系统进行交互的自然人告知该事实。聊天机器人是代表性的例子。要进行公开，以便使用者认识到自己正在与机器进行对话。这与韩国法律的事先告知义务方向相同。

在深度伪造内容上必须贴上由AI生成或操纵的标签。这对应了韩国法律的生成物标识义务。

第三个范畴是韩国法律中没有的。对为了向大众告知公共关注事项而公开的AI生成文本也赋予了标识义务。这是针对新闻报道或用于公共讨论之AI生成文本的条款。其意图是为了防止AI撰写的报道与人撰写的报道毫无区别地流通的情况。不过，如果是经过了人的审查和编辑控制，且由自然人或法人承担编辑责任的情况，则免除标识义务。韩国《人工智能基本法》虽然也对生成式AI撰写的文章适用标识义务（第31条第2款），但并没有专门针对公共关注事项文本的范畴。如何处理新闻领域的AI标识，仍然是作为单独讨论的课题而遗留下来的。

欧盟对通用AI（GPAI）模型的透明度义务更进了一步。GPAI行为准则（Code of Practice，2025年8月适合性决定）要求模型提供者将技术文档保存10年。必须将学习数据来源摘要、版权政策（尊重robots.txt）和系统性风险管理方案文档化。韩国法律中没有与之对应的文档保存期限规定。

欧盟的综合修正案（2026年6月29日欧盟理事会最终批准）增加了水印宽限条款。对2026年8月2日之前推出的系统给予4个月的宽限。这是为了给现有服务适应新标识义务的时间。虽然这与韩国法律实施后设定1年的指导期的主旨相似，但欧盟仅对特定义务给予了短暂的宽限，而韩国则对整个行政罚款给予了广泛的宽限。

中国选择了不同的路径。2025年9月1日实施的《人工智能生成合成内容标识办法》要求进行双重标识。必须同时贴上显性标识和隐性标识。显性标识是使用者肉眼可见的标签。即“此内容由AI生成”的提示。隐性标识是在元数据或数字水印中植入机器可读的信息。甚至还制定了强制性国家技术标准（GB 45438-2025），将标识的技术规格标准化。与韩国法律仅对深度伪造性结果物要求明确的认识方式相比，中国是在AI生成内容的整体上贯彻了双重标识。

韩国广泛地认可例外。如果是内部业务用就免除，如果服务名称很明显是AI就免除。欧盟则细分了范畴且涵盖了公共文本。中国则缩小了例外并推行技术标准。如果将透明度的广度按顺序排列，大致是按照中国、欧盟、韩国的顺序。韩国的方法蕴含了通过减轻监管负担来帮助产业初期成长的判断。这种判断是否能充分保护使用者的知情权，还有待于法律实施后积累案例来进行验证。

8. 如果违反，会发生什么变化

违反了透明度义务，并不会立即被处以行政罚款。行政罚款的对象是有限的，而且还有指导期。

《人工智能基本法》第43条是处以行政罚款的依据。上限为3000万韩元。但是，在第43条列举的行政罚款事由中，直接涉及透明度的只有违反事先告知（第31

条第1款)。违反生成物标识(第31条第2款、第3款)本身并不是行政处罚的对象,只有在经过事实调查和纠正命令后仍不履行命令时,才会导致行政处罚。这其中还重叠了指导期。科学技术信息通信部表示,在法律实施后至少1年、到2027年1月以后为止,将作为指导期进行运作。在此期间如果发现遗漏透明度义务或文档化不完善,将优先适用行政指导和纠正建议,而不是行政处罚。其主旨是给经营者自我改正的机会。

即使是指定期,也并不意味着所有违规行为都会被免责。在诸如利用深度伪造进行的犯罪、由AI系统造成的人身事故以及大规模侵犯人权等重大事件中,不论处于什么期间,都可以立即进行事实调查。指导期是为了保护善意准备不足的人,而不是容忍恶意违反或造成重大损害。

指导期结束后的3000万韩元行政处罚威慑力很弱。在《欧盟AI法案》中,违反透明度义务(第50条)的制裁上限是1500万欧元(约240亿韩元)或全球年销售额的3%。被禁止的AI惯例另外设有3500万欧元或7%的更高上限。韩国的3000万韩元与1500万欧元之间有着将近800倍的差距。中国则根据违规类型准用相关法律进行处罚,在行政处罚之余还可能导致停业和停止服务。对于年销售额达到数千亿韩元的大企业来说,3000万韩元连一年的备品购买费都不到。在违法行政处罚比守法成本更便宜的结构中,行政处罚很难发挥制裁作用。

但是,行政处罚并不是唯一的压力。如果发现违规事实或接到举报,科学技术信息通信部可以进行事实调查(法律第40条第1款),如果调查结果认定存在违规,就会下达中止命令或纠正命令(第40条第3款)。虽然法律中没有另外设立调查结果公布制度,但如果“这家公司的AI服务没有向使用者告知AI使用事实”这一事实被传开,对于以使用者信任为业务基石的AI服务来说,将是比行政处罚更大的打击。违规事实暴露本身就有可能成为事实上的威慑力。

在指导期结束的2027年以后,科学技术信息通信部采取事实调查、纠正命令和处以行政处罚的积极程度,将决定该法律的实效性。条款再怎么精妙,如果缺乏

执行，也就只能停留在建议层面。韩国的3000万韩元即使与欧盟的1500万欧元相比，也只是象征性的水平，因此需要通过执行的积极性来弥补行政罚款金额上的弱点。

9. 留待解答的问题

仍然存在透明度指南尚未给出具体答案的领域。

AI生成的代码也必须贴上标识吗？当开发人员将用GitHub Copilot制作的代码包含在软件产品中时，是否必须在该软件上标注“包含AI生成代码”。代码是否属于法律所列举的“文章、声音、图画、视频”，这一点从一开始就不明确。

如果AI写了初稿，人在修改了70%后公开，必须在结果物上做AI标识吗？AI贡献比例超过哪条线才成为标识对象，指南中并没有标准。如果人的修改足够多，它是否就成为人的著作物，还是只要AI参与了初稿就需要标识？这是透明度和版权的交叉点，而在任何一方的指南中都未被梳理出来。

这是人最终批准AI所推荐之判断的情况。AI贷款审查工具推荐了“批准”，顾问在确认后发放了贷款。在透明度告知中，该决定必须标记为“AI做出的判断”，还是必须标记为“由AI辅助之人的判断”？如果认为是人按下了最终按钮所以是人的决定，告知义务就会减轻；如果认为是AI实质上得出了结论所以是AI的判断，告知义务就会加重。

AI基本法支持台的电话响个不停。在552件咨询中大约有一半是透明度问题的数字表明，法律和指南虽然确立了原则，但实务上的空白还很多。这个空白只能由经营者和监管机构通过不断积累案例来填补，而支持台受理的每一个咨询都是该项工作的材料。

第四部 强化义务对象：高性能与高影响人工智能

第10章. 高性能人工智能经营者的确保安全性义务

2026年2月，在首尔江南的一家AI初创企业会议室里，一名工程师敲打着计算器。他正在将在自身语言模型的学习中投入的累计运算量换算为FLOP。数字出来了。10的23次方浮点运算。这与《人工智能基本法》施行令规定的基准线即10的26次方相差1000倍。并非只有这家公司如此。截至2026年7月，尚未发现在韩国国内拥有跨过该门槛模型的经营者的。第32条所瞄准的适用对象，是像OpenAI、Google DeepMind、Anthropic、Meta这样将数十亿美元倾注于学习基础设施的极少数前沿研究所。然而，该法律却由韩国国会刊登在韩国公报上。这种张力便是解读第32条的出发点。

1. 第32条的结构

《人工智能基本法》第32条由三个款组成。第1款规定了经营者应当遵守的实体义务，第2款附加了提交其履行结果的义务，第3款则将制定详细标准的事宜委托给科学技术信息通信部长官的公告。

第1款第1项要求“在人工智能生命周期全过程中的风险识别、评估及缓解”。“生命周期”一词是核心。其含义是在从模型设计阶段到学习、评估、部署、运营、废弃的整个过程中，必须无一遗漏地进行管理。找出风险（识别），衡量该风险有多大（评估），并采取减少风险的措施（缓解）。这三个阶段的程序被压缩在法条之中。

第1款第2项是“构建监控并应对人工智能相关安全事故的风险管理体系”。这指示在将模型推向世界后也不要停止监视。事后监控已上升为法律上的义务。

第2款要求将履行这两项义务的结果提交给科学技术信息通信部长官。并非自主完成就结束了。这附加了向政府报告的程序。

第3款委托长官通过公告规定履行方式和提交程序。基于该委托的公告截至2026年7月已结束行政预告，制定程序正在进行中。

一个条文汇集了义务（第1款）、报告（第2款）和详细标准委托（第3款）。结构很简洁。问题在于，这简洁的句子究竟要求实际的经营者采取何种行动。

2. 施行令第24条：三个要件

第32条第1款仅写道“学习中使用的累计运算量达到总统令规定标准以上的人工智能系统”。这便是将基准线本身交由施行令规定。

施行令第24条第1款规定，必须同时满足三个要件。

第一。学习中使用的累计运算量应达到 10^{26} 次方FLOP（浮点运算）以上。

第二。应是应用当前人工智能技术中最前沿技术来构成和运营的。

第三。系统的风险度存在对人的生命、身体的安全及基本权利造成广泛且重大影响的担忧。

这三个要件是累加（AND）条件。即使运算量超过标准，若非最前沿技术，则不适用。即使是最前沿技术，若风险不广泛，也同样不适用。

让我们暂时观察一下 10^{26} 这个数字。OpenAI在学习GPT-4时投入的运算量，外部估算值约为 2×10^{25} 次方FLOP。如果该估算准确，GPT-4也未能达到韩国法律的标准。超过基准线的模型，只有在落后GPT-4一代以上的时期，也就是GPT-5级别或Google Gemini Ultra的后续模型、Anthropic的Claude Opus 4级别的最顶层模型中才会出现。这比欧盟《人工智能法案》（EU AI Act）设定的通用人工智能（GPAI）模型系统性风险分类标准—— 10^{25} 次方FLOP高出一个数量级，即10倍。韩国的门槛是被刻意定高的。

施行令第24条第2款规定，累计运算量的具体计算方式由科学技术信息通信部长官通过公告规定。是否要将模型学习以外的微调（fine-tuning）、强化学习阶段的运算加总，混合精度（mixed precision）运算的换算比例该如何设定等技术性问题，将在该公告中得到解决。

3. 生命周期全过程的风险管理

法条所说的“生命周期”是指AI系统的整个生命过程。即从设计阶段筛选学习数据开始，历经模型学习、评估、部署、服务运营、更新，直至服务终止。

风险识别是在各个阶段编制可能发生何种有害结果之清单的工作。学习数据中是否混有偏见，从而得出对特定种族或性别不利的结果？模型是否会生成生物武器合成法或网络攻击代码？使用者是否会操纵模型的输出，从而大量散布虚假信息？模型是否会自行生成下级智能体，从而脱离控制？清单的范围很广。

评估是对已识别出的各个风险的发生可能性和严重程度进行量化或分级的过程。也就是要衡量“该模型提供CBRN（化学、生物、放射性、核）相关知识的概率是多少，该知识导致实际损害的可能性有多大”。

缓解是根据评估结果采取减少风险措施的阶段。从学习数据中去除有害内容、安装输出过滤器或将模型策略设定为拒绝特定类型的请求等，均属于此列。将风险降至可容许的水平，而非将风险降为0，才是缓解的目标。

4. 安全事故监控体系

第32条第1款第2项要求的监控，是指在部署模型后也要继续进行监视。在实际业务中，通常会通过实时监视体系来实现。

前沿模型驱动着数亿人同时使用的服务。在学习过程中未能发现的漏洞，可能会在实际使用过程中暴露出来。2023年3月，OpenAI发现其他使用者对话列表的标题、新开始对话的第一条消息的一部分以及ChatGPT Plus部分订阅者的支付信息被泄露的错误，随后暂时停止了服务。2024年，Google的Gemini在生成图像的过程中，因对历史人物种族的描绘不准确，曾出现过暂停服务的案例。这类事故仅凭事前评估是无法预防的。

监控体系由三个支柱构成。即异常检测（实时追踪模型输出的有害性、偏见和错误率的系统）、事故报告（使用者举报渠道和内部报告程序）以及应对程序（

发生事故时采取分析原因、停止服务或发布补丁等措施的流程)。经过行政预告的公告草案包含了记录并保存监控结果，并在发生安全事故时向科学技术信息通信部长官报告的内容。

5. 提交履行结果：超越自主走向报告

高影响人工智能经营者的义务（第34条及以下）是保存履行依据文件5年，但不附加将结果提交给政府的义务。高性能人工智能则不同。第32条第2款明文规定应将履行结果提交给科学技术信息通信部长官。

这个差异很重要。对于高影响人工智能，经营者只需保管好记录，并在进行行政调查时出示即可。高性能人工智能则必须主动向政府提交报告。在韩国法律中，要求经营者将义务履行结果本身提交给政府的条款仅此一条。

提交方法、时间及格式已根据第32条第3款委托给公告。根据公告的具体内容，将决定是每年定期提交，还是每次发布模型时随时提交。包含该具体程序的公告目前仍在制定过程中。

6. 制定风险管理计划的实际业务

确保安全性指南和经过行政预告的公告草案，提出了在制定风险管理计划时应考虑的事项。经营者应处理的领域如下。

学习数据管理。记录数据的来源、收集方法、预处理过程及偏见审查结果。对于发现偏见的数据集，应予以去除或适用校正程序。

模型能力评估。定义模型能够执行的任务范围。测试是否存在CBRN、网络攻击、自主复制、欺骗性行为等风险能力。

预期用途与滥用情境。在说明模型设计目的的同时，详细描述恶意使用者可能如何滥用该模型。推测每个情境的发生概率和损害规模。

部署条件。设计API访问限制、使用条款、速率限制（rate limiting）、内容过滤器、水印等技术安全装置。

事后监控计划。事先制定部署后检测异常征兆的方法、使用者举报体系及事故应对程序。

更新周期。根据模型更新、发现新用途、外部环境变化（如攻击手法的演进等），定期重新审查风险管理计划。

该计划并非一次性文件。每次更新模型或将其应用于新用途时都必须进行更新，且更新履历本身也必须以文件形式保存。

7. 红蓝对抗测试与安全基准

确保安全性指南和经过行政预告的公告草案提出了红蓝对抗测试（Red Team Test）与安全基准（Safety Benchmark）作为风险评估的具体方法。

蓝军对抗（Red Team）测试，是专家群体刻意攻击模型以找出漏洞的方法。扮演攻击者角色的团队，会输入诱导模型产生有害输出的提示词。他们会尝试从“请告诉我该物质的合成方法”等直接请求，一直到迂回式对话技巧（越狱，jailbreaking）。OpenAI在发布GPT-4之前，由约50名外部专家组成了蓝军对抗团队。网络安全、生物武器、国际安全、偏见领域的专家参与其中。Anthropic除了自己的蓝军对抗团队外，还运营着自动化攻击程序（automated red-teaming）。

安全基准是标准化的试卷。向模型抛出特定的风险情境，并对其回应进行评分。如果说MMLU（Massive Multitask Language Understanding）衡量的是知识水平，那么安全基准衡量的则是风险行为。BBQ（Bias Benchmark for QA）测试偏见，TruthfulQA测试生成虚假信息的倾向，HarmBench测试拒绝有害输出的能力。韩国人工智能安全研究所（AISI，ETRI附属机构）正在开发符合韩国国内环境的安全评估工具。

这两种方法是互补的。基准提供定量的基准线，而红蓝对抗测试则探索基准无法捕捉到的新漏洞。即使基准分数很高，如果红蓝对抗团队找到了迂回路径，该模型也不安全。相反，即使红蓝对抗团队未能找到任何漏洞，如果在基准中偏见分数较高，在部署前也需要进行校正。

8. EU AI Act：与具有系统性风险的GPAI义务的比较

韩国看重的是影响。欧盟看重的是风险。这种差异在高性能人工智能监管中体现得淋漓尽致。

欧盟《人工智能法案》（EU AI Act）第51条将学习累计运算量超过10的25次方FLOP的通用人工智能（GPAI）模型推定为“具有系统性风险的GPAI”。这是只要经营者提出反面证据即可推翻的推定。这比韩国施行令第24条的基准（10的26次方）低一个数量级。欧盟的网撒得更广。

欧盟第55条对这一级别的GPAI提供者赋予了四项义务。执行包含对抗性测试在内的模型评估。评估并缓解系统性风险。追踪、记录重大事故，并向AI办公室及各国监管机构报告。确保模型的网络安全。

与韩国第32条有重叠的内容。风险识别、评估、缓解以及事故监控，对双方都是共通的。差异在三个地方显现出来。

是网络安全义务。欧盟将GPAI模型本身的网络安全明文规定为一项单独义务。韩国法律中没有这一条款。网络安全虽有被囊括为风险管理其中一项的余地，但并未被分离为独立义务。

是行为规范（Code of Practice）。欧盟第56条允许经营者通过自愿的行为规范来履行义务。2025年7月，约1000名利益相关者参与并确定了最终规范，8月批准了合格判定。该规范由透明度（文件化，保存10年）、版权（尊重robots.txt，将侵权降至最低）以及安全与保安（系统性风险管理）等三个章节组成。该规范将CBRN能力、失控、网络攻击、有害操纵及欺骗列为四大系统性

风险。韩国目前尚无与此对应的民间规范体系。

是运算量基准线的差距。欧盟的10的25次方能捕捉到GPT-4级别的模型。韩国的10的26次方则会漏掉GPT-4。这一个数量级的差距改变了监管对象的范围。以欧盟标准能捕捉到数十个前沿模型，但以韩国标准，其中只有一部分会被纳入。

制裁水平也不同。对于违反具有系统性风险的GPAI规定的行为，欧盟可处以1500万欧元或全球年营业额3%的罚金，而韩国《人工智能基本法》的行政罚款上限为3000万韩元。相差了500倍以上。不过，在2026年8月2日，第50条透明度义务和能够处以该罚款的GPAI执行权将按原计划适用，而高风险AI的义务则通过综合修正案推迟至2027年12月2日。

9. NIST AI RMF：与自愿性框架的比较

美国联邦层面上并没有针对高性能人工智能的法律上的义务。拜登政府曾于2023年10月颁布第14110号行政命令，要求前沿模型开发者向政府提交安全性评估结果，但在2025年1月，特朗普政府撤销了该命令。联邦层面的强制义务消失了。而在州层面则有所行动。加利福尼亚州于2025年9月制定了《前沿人工智能透明度法》（SB 53），要求以超过10的26次方FLOP的运算进行学习的模型的开发者必须公开安全框架并报告重大安全事故，该法自2026年1月起实施。

留在联邦层面的，只有NIST（美国国家标准与技术研究院）的AI风险管理框架（AI RMF）。这项于2023年1月发布的框架，是没有法律约束力的自愿性指南。

AI RMF由四项核心功能构成。即GOVERN（建立治理结构）、MAP（掌握风险脉络）、MEASURE（风险衡量）和MANAGE（风险管理）。韩国法律第32条在法律条文中压缩为“识别、评估、缓解”的内容，NIST则将其拆分为MAP、MEASURE、MANAGE这三项功能。GOVERN指的是组织层面的管理体系，与韩国法律中的构建风险管理体系相对应。

2024年7月，补充了针对生成式人工智能的资料（NIST AI 600-1）。该文件将生成式人工智能固有的12种风险，分类为CBRN、充满确信的虚假生成（confabulation）、偏见、网络攻击、隐私、环境影响等。

将韩国、欧盟、美国的方法并列起来看，便能看到鲜明的光谱。欧盟通过法律义务与高额罚金进行强制，并附加了名为行为规范的民间履行手段。韩国将法律义务与政府报告相结合，但将行政罚款设定得较低。美国在联邦层面没有法律义务，交由自愿性框架及业界的自我监管，但像加利福尼亚州一样开始用法律介入的州正在涌现。运算量基准线方面，欧盟是10的25次方，韩国和加利福尼亚州是10的26次方，而美国联邦根本不存在基准。

10. 极少数的适用对象

截至2026年7月，尚未发现在韩国国内学习超过10的26次方FLOP模型的经营者。Naver的HyperCLOVA X、三星的Samsung Gauss、LG的EXAONE、KT的Mi:dm（信仰）等国内大型模型，是数百亿参数的规模，其学习运算量估算在10的23次方到24次方的水平。与基准线的差距达100倍到1000倍。

第32条的实质适用对象是海外前沿研究所。根据《人工智能基本法》第4条第1款（域外适用），“即使是在韩国境外实施的行为，若对韩国国内市场或使用者产生影响的”，他们便成为适用对象。这与指定韩国境内代理人的义务（第36条）相咬合，构成了一种只要海外前沿研究所在韩国提供服务就无法摆脱此项义务的结构。

那么，这项条款对韩国国内经营者来说就没有意义了吗。

基准线并非是固定不变的。运算量计算方式由科学技术信息通信部长官根据施行令第24条第2款通过公告规定，而要改变10的26次方这一基准线本身，则必须经过施行令修改程序。用于相同规模学习的费用每年都在快速下降。GPU性能每一代都在提升，学习算法的效率也在变好。韩国国内经营者达到10的26次方的时间点，可能会在3年到5年内到来。只是，并非一旦达到运算量标准就立即成为适

用对象。必须同时满足应用最前沿技术以及具有广泛且重大风险这两个剩余要件，才适用施行令第24条。

法律传递出信号。即使当下不适用，提前具备风险管理体系的企业，在日后就能减少监管转换的费用。这便是为何确保安全性指南不仅适用于高性能人工智能经营者，还作为所有AI经营者的参考资料予以公开的原因。

域外适用与韩国境内代理人制度要求海外前沿研究所在韩国市场遵守行为规范。当韩国使用者使用OpenAI或Google的模型时，该模型是否满足韩国法律要求的安全性标准，科学技术信息通信部将能够通过提交的履行结果予以确认。这一程序的实效性尚未得到验证。

江南会议室的工程师放下了计算器。10的26次方是一个遥远的数字。AI学习运算量的增长曲线每年呈现出接近4倍的斜率，而费用曲线正朝着相反的方向下降。在这两条曲线相交的某个地方，这项条款才会真正鲜活地运作起来。

第11章. 高影响人工智能经营者的五项责任

首尔的一家金融科技公司使用AI信用评估模型审查小微企业主的贷款申请。2026年2月，AI拒绝了一位经营餐厅7年的店主的贷款申请。店主问道：“为什么拒绝？”金融科技公司的法务团队翻开了《人工智能基本法》。第34条第1款第2项要求，“在技术可行的范围内”解释AI得出结果的主要标准。神经网络模型参数有2亿个。“技术可行的范围”到哪里为止，法务团队也没有答案。解释是五项责任之一。风险管理、使用者保护、人的监督、文件保存是连同一起运转的。

《人工智能基本法》将提供高影响人工智能的经营者的责任集中在第34条这一个条文中。第34条第1款并列列举了五项措施。风险管理方案的制定与运营（第1项）。解释方案的制定与实施（第2项）。使用者保护方案的制定与运营（第3项）。人的管理与监督（第4项）。履行依据文件的制作与保存（第5项）。由于后面还附有包含国家人工智能委员会审议并表决事项的第6项，该列表是开放的。这并不是五项独立的义务。而是一个循环体系。管理风险，解释其结果，保护使用者，让人能够进行控制，并记录所有过程。如果缺少其中一项，其余的将失去意义。

1. 管理风险意味着什么

提供高影响人工智能的经营者应当具备的首项责任是风险管理方案的制定与运营。

法律并不将风险管理视为一次性的检查，而是一个循环的过程。科学技术信息通信部发布的经营者责任指南将这一过程分解为五个阶段。政策制定。风险识别。分析与评估。处理。改进。

政策制定是指确定风险管理的范围与责任。以书面形式确定哪些AI系统是对象，由谁总管管理，以及报告路径在哪里。如果是运营贷款审查AI的金融科技公司，将该系统分类为高影响、指定风险管理负责人并确立报告体系就是起点。

风险识别是找出该系统可能对使用者及受影响者造成的危害。如果是贷款审查AI，针对特定年龄段或地区的系统性偏见就是风险候选。如果是医疗影像判读AI，在特定人种肤色上准确度下降的现象就是候选。如果是招聘筛选AI，倾向于学历或性别的模式就是候选。

分析与评估是衡量已识别风险的严重程度和发生可能性。在1万件贷款审查中，因偏见被不当拒绝的案件是10件还是100件，该拒绝给个人生计带来的影响达到何种水平，要进行定量估算。对于无法定量的风险，则并行开展定性评估。

处理是执行降低风险措施的过程。一旦确认偏见，就重构训练数据，调整决策边界，或将特定变量从模型中排除。如果无法完全消除风险，则判断剩余风险是否处于可接受的水平。如果无法接受，就必须停止系统运营。

改进是回顾风险管理结果并修改政策的过程。确定周期（无论按季度还是半年度）检查成果，发现新风险时回到识别阶段。每当模型更新、训练数据变更、服务范围扩大时，这个循环就会重新运转。

施行令第27条第1款第1项要求将风险管理政策与组织体系的主要内容发布在营业场所或官方网站上。属于商业秘密的内容可以排除，但必须公开风险管理方案存在的事实及其骨架。这种结构让使用者能够确认经营者意识到了哪些风险，以及由谁承担管理责任。

欧盟AI法第9条对同一主题进行了详尽得多的规定。法律文本本身包含“贯穿高风险AI系统整个生命周期进行计划和实施的持续和迭代过程（continuous iterative process）”的定义。法条中直接明文规定了四个阶段。识别和分析已知风险及合理可预见的风险。在预期使用及合理可预见的误用情况下的风险估算与评估。基于上市后监测数据评估额外风险。采取适当且有针对性的风险管理措施。

第9条第5款甚至提出了处理剩余风险的三条路径。通过设计消除或减少。对无法消除的风险采取缓解和控制措施。向用户提供信息和培训。第9条第9款要求单独考虑对18岁以下儿童和弱势群体的影响。

韩国法律与欧盟法律的区别在于约束力。欧盟第9条是直接适用的硬性规定（regulation）。一旦违反，最高将被处以1500万欧元或全球营业额3%的行政罚款。韩国第34条第1款第1项的风险管理义务将具体履行方法委任给施行令与公告，并规定科学技术信息通信部长官“可以建议予以遵守”（第34条第2款）。即使不履行，也不会立刻被处以行政罚款。经过事实调查后下达停止命令或纠正命令，且在不履行该命令时，才会被处以最高3000万韩元的行政罚款（第40条、第43条第1款第3项）。政府已表示将实施至少1年的指导期。

中国没有设置统一的风险管理义务。而是要求根据服务类型进行安全评估。2023年8月施行的《生成式人工智能服务管理暂行办法》对具有“舆论属性或者社会动员能力”的服务附加了事前安全评估义务。这是一种比宽泛的风险管理框架更具体的、以内容安全为轴心的处理方式。

2. “为什么会出现这种结果”

第34条第1款第2项要求制定并实施解释方案。法律要求解释的对象有三项。人工智能得出的最终结果。得出该结果所使用的主要标准。开发所使用的训练数据概要。

施行令第27条第1款要求将解释方案的主要内容和使用者保护方案发布在官方网站上。还公开负责管理的人员的姓名与联系方式。可以排除商业秘密，但方案本身的存在不能隐瞒。

“在技术上和合理上可行的范围内提供明确且有意义的解释”这一措辞源自基本原则第3条第2款。这是一个宣布受影响者应该能够获得明确且有意义的解释的原则的条款。这句话中包含了两个争议点。

“技术可行的范围”到哪里为止。在当前技术水平下，完全解释拥有数亿参数的大规模语言模型等深度学习模型的内部运作原理是不可能的。法律认识到了这一点。附加上“在可能的范围内”的条件，意味着不要求解释的完整性。那么，经营者至少应该做些什么。

指南指明了方向。对于最终结果，以使用者能够理解的形式告知AI做出了何种判断。如果是贷款审查，结果就是“您的申请未获批准”。关于主要标准，解释对结果产生影响的高层因素。也就是信用等级、收入负债比例、最近6个月内的交易记录产生了负面作用等程度。关于训练数据概要，告知使用了哪种类型的数据进行训练的范畴和规模。也就是“使用了5年间积累的100万件贷款审查记录进行训练”这种水平。

并不是要求公开模型的权重或源代码。商业秘密受到明文保护。经营者要做的是，提供使用者能够接受针对自己所下判断之背景的程度的解释。

剩下的是“有意义的解释”的标准。列出模型计算出的特征重要性（feature importance）分数并说“已提供解释”，在技术上是可行的。然而，收到满是变量名称和数值的报告的贷款被拒者能否理解它，这是另外一个问题。“有意义的”这个修饰语，是要求将解释的水平对准受影响的人的眼界，而不是专家。这不是技术文件，而是沟通。

使用者保护方案是独立于解释方案之外的责任。第34条第1款第3项单独要求制定并运营使用者保护方案，施行令则要求将该方案与解释方案一起发布在官方网站上。虽然法律和施行令没有直接规定内容列表，但其中包含异议提出程序、损失救济路径、负责部门联系方式等内容。必须明确，对AI判断有异议的使用者可以向哪里、以何种方式、在多少时间内提出异议。

欧盟AI法第13条将其命名为“透明度与向部署者提供信息”并加以规定。第13条第3款列举了使用说明书应包含的内容。提供者的身份和联系方式。系统的特性、能力和性能限制。预期目的和准确度指标。鲁棒性与网络安全水平。可预见

的风险。可解释性（explainability）。特定群体的性能差异。输入数据规格。解释指南。人的监督措施。计算资源与预期寿命。这是十二项以上的清单。

欧盟将清单直接写进了法条。经营者按照清单执行即可。韩国只写了“制定并实施解释方案”，其余交给了公告和指南。欧盟的方法赋予了经营者可预测性。做到什么程度才算合格是明确的。韩国的方法是灵活的。由于各行业背景不同，这是不通过法律固定统一清单的判断。而另一方面，经营者必须自己决定“做到什么程度才算足够”。

这种负担是2026年7月目前的实务景象。在AI基本法支援服务台实施10周内受理的552起咨询中，51%是关于透明度标识义务的提问（科学技术信息通信部2026年3月31日发布）。经营者们正在询问解释的范围和深度。

3. 人手可及的距离

第34条第1款第4项要求对高影响人工智能进行人的管理与监督。施行令要求将负责管理与监督的人员的姓名与联系方式在官方网站上公开。

法律的意图是这样的。在高影响领域，人应当管理并监督AI，建立必要时可以介入的体系。

要将这一点落实到实务中，必须具备三项机制。

介入权限。必须有人能够推翻AI的结果。当贷款审查AI得出拒绝结果时，负责员工必须拥有推翻该拒绝的权限与程序。仅仅事后确认AI的判断是不够的。确认后必须能够更改。当招聘筛选AI淘汰了申请人时，人事负责人必须能够无视（override）该决定，将其提上进入面试的名单。

紧急停止。当AI系统运行异常时，人必须能够停止系统。如果医疗影像判读AI开始将正常组织判读为肿瘤，必须设有医生或系统运营者可以立即中止该AI运行的程序。如果是自动驾驶辅助系统，驾驶员可以转换为手动模式的物理界面就属于紧急停止。紧急停止可以是软件按钮，也可以是物理开关，或者是利用管理员

权限远程阻断的程序。形式依系统而异，但人能立即停止AI这一事实必须是一致的。

介入触发器。必须事先规定人在什么情况下介入。可以制定规则，若贷款审查AI在特定期间的拒绝率超过平时的两倍，就由人介入。可以制定规则，若招聘筛选AI将特定大学出身的申请人淘汰90%以上，就由人进行审查。可以制定规则，若能源供应管理AI记录了15%以上的预测误差率，就由运营者转换为手动模式。触发器最好用数值定义。“发现异常就介入”这样模糊的标准，实际上等同于对人的介入放任不管。

将负责管理与监督的人员姓名与联系方式公开在官方网站上的条款，其意在将责任绑定于特定人。如果AI出现问题，由谁来响应，使用者及受影响者必须能够知道。此人必须理解AI系统的运行，能够掌握异常征兆，并拥有停止系统的权限。

欧盟AI法第14条更加细致地规定了人的监督（human oversight）。列举了监督负责人员应当具备的五项能力。理解系统的能力与局限性。认识自动化偏见（automation bias）的风险。正确解释系统的输出。能够拒绝或无视系统的输出。能够介入或安全停止系统。

自动化偏见这个概念很引人注目。它是指人过度依赖AI的判断而放弃自身判断的现象。如果监督者坐在那个位置，却只是照单批准AI得出的结果，那么人的监督就流于形式。欧盟第14条明文规定认识自动化偏见的风险，就是为了警惕这种形式化。韩国法律中没有关于自动化偏见的提及。

第14条第5款对生物识别系统设定了单独的标准。如果远程生物识别系统识别了个人，除非至少有两名自然人分别进行验证和确认，否则不得采取任何措施或决定。韩国法律的高影响领域中，“为犯罪调查或逮捕而分析生物识别信息”（第2条第4项F目）涉及了与欧盟此项规定相同的领域，但至少两人交叉验证这种具体的要件，在韩国法律中并不存在。

中国学术界2024年推出的综合人工智能法学者建议稿（共9章96条）在影响合法权利的决定中包含了人的介入（human-in-the-loop）要求。其中提及了最高5000万元人民币（约90亿韩元）的罚款。已经施行的《生成式人工智能服务管理暂行办法》虽然以事前安全评估的形式要求了人的验证，但与个别决策单位的人的监督在性质上有所不同。

4. 记录并留存的5年

第34条第1款第5项要求制作并保存履行依据文件。施行令第27条第2款将保存期限规定为5年。也允许采用电子方式。

保存对象是在五项责任履行过程中制作的全部文件。风险管理方案的制定与变更记录。风险识别与评估结果。解释方案的内容与使用者保护措施。人的介入体系的运营记录。当进行事实调查、使用者提出异议或发生事故时，这些文件就成为证明经营者是否履行责任的依据。

5年的期限反映了一种判断，即高影响人工智能的影响可能会随着时间差显现出来。招聘筛选AI的偏见在现实中显现出来可能需要数年时间。贷款审查AI的系统性歧视若要在统计中被捕捉到，必须积累足够的数据。公共服务AI的资格审查错误被积累并被认知为社会问题之前，会流逝相当长的时间。

在经营者看来，5年保存是一项成本。在AI系统频繁更新的环境下，要将所有版本的风险管理记录和解释方案维持5年，必须另行建立文件管理体系。如果是模型以月为单位重新训练的服务，5年间积累的文件量是不容忽视的。

欧盟AI法将文件编制和记录分成两个条文。第11条（技术文件）要求在投放市场前编制技术文件并保持更新。最低包含事项列举在附件IV中，对于中小企业，允许使用简化的文件形式。第12条（记录保存）要求高风险AI系统具备自动记录事件的技术能力。这就是自动记录（automatic recording of events）。导致风险的情况识别、上市后监测、与运营监视相关的事件属于记录对象。

韩国法律的文件编制义务将焦点放在人制作的文件上。欧盟的记录保存义务甚至涵盖了系统自行留下的日志。欧盟第12条要求的自动记录是使事后审计（audit）在技术上成为可能的机制。韩国法律中没有对应自动记录的规定。这意味着法律并不要求具备可以机械追踪AI系统个别决策原因的体系。

保存期限也存在差异。以自愿签署方式运作的《GPAI行为准则》（2025年8月批准）要求已签署的通用AI提供者保存技术文件10年。与韩国的5年相比是两倍。欧盟AI法本身将技术文件的保存期限规定为系统投放市场后10年。韩国的5年是欧盟的一半。

5. 一个条文与七个条文

韩国《人工智能基本法》将高影响人工智能经营者的责任集中在第34条这一个条文中。第1款的五项内容列举了风险管理、解释、使用者保护、人的管理与监督、文件的制作与保存。欧盟AI法利用七个条文（第9条至第15条）规定了高风险AI系统的义务。

结构造就密度。欧盟AI法第9条至第15条加起来，有数十项详细要求直接列在法条中。韩国仅在法律中写下五项大责任，细节事项则下放给施行令、公告和指南。对于欧盟的规定，经营者只需阅读法典就知道该做什么。韩国的基本法仅靠法典是不够的，必须经过施行令、公告、指南甚至支援服务台咨询，才能把握全貌。

有两项义务是欧盟有而韩国没有的。

数据治理。欧盟第10条用单独的条文处理用于训练、验证、测试的数据集的质量标准。列举了数据收集过程和来源、预处理操作（标注、标签、清洗、增强）、偏见审查、代表性确认以及个人信息保护特例等八项内容。韩国《人工智能基本法》中没有直接规制训练数据质量或治理的条款。虽有要求解释训练数据“概要”的规定，却漏掉了建立数据本身质量标准的规定。

准确性、鲁棒性和网络安全。欧盟第15条要求高风险AI系统达到适当水平的准确度、鲁棒性（robustness）和网络安全。从对错误和缺陷的恢复力、对数据中毒（data poisoning）和对抗性示例（adversarial examples）的防御，到防止因反馈循环扩大偏见，均在法条中明文规定。韩国法律中没有对应的独立条款。虽然可以作为风险管理的一环被包容进去，但在法律层面明文规定与在公告层面处理之间，约束力是不同的。

对韩国法律这一结构属于“灵活”的评价是准确的。在AI技术快速变化的状况下，如果将细部标准固定在法律中，那么每次技术跑在前面时，就必须修改法律。公告和指南无需修法即可更改。但是，存在着灵活性转化为不确定性的节点。如果法律本身不告知应做的事，那么有守法意愿的经营者也不知道该遵守什么。这552起提问就是这种不确定性的一个侧面。

6. 在建议与义务之间

第34条第1款的五项责任中附加了“应当履行”的措辞。这是义务。然而，对于科学技术信息通信部长官在公告中规定的具体事项，却写着“可以建议予以遵守”（第34条第2款）。这是建议。

在义务与建议之间存在间隙。必须履行五项责任是法律上的义务，但如何履行的详细标准则是建议。围绕这种结构，使用了“自主性努力义务”这种表达。

行政罚款最高为3000万韩元。而且不是违反了责任就立刻被处以罚款，而是在经过事实调查后下达的停止命令或纠正命令未被履行时，才会被处以罚款（第40条第3款、第43条第1款第3项）。政府公布的指导期至少为1年，将持续到2027年1月之后。除非是人员伤亡或大规模人权侵犯等重大事由，否则会先进行行政指导和纠正建议，而不是行政罚款。

3000万韩元这个金额对有10名员工的初创企业来说是负担。对年销售额1万亿韩元的大企业来说仅是0.003%。有比较对象。欧盟AI法的最高风险AI义务违反罚款最高为1500万欧元或全球营业额的3%。将1500万欧元换算成韩元约为230亿

韩元，与韩国的最高行政罚款相差近800倍。中国学者建议稿提出的罚款最高为5000万元人民币（约90亿韩元）。韩国的3000万韩元在三国之中显著偏低。

罚款的数额并不等于规制的实效性。韩国法律通过其他途径引导经营者的行为。

就是公共采购联动。基本法第35条第2款规定，国家机关如果想使用高影响人工智能产品或服务，应优先考虑实施了影响评估的产品。2026年1月修订时加入的第16条第3款规定，国家机关在购买业务所需的产品或服务时，应优先考虑人工智能产品和人工智能服务。该修订条款自2026年7月21日起施行。将两项条款合在一起就成了这样的结构。对于想将政府作为客户的经营者来说，履行影响评估和履行责任实际上是市场准入条件。法律没有强制的，市场会强制。如果政府采购市场不大，这种联动就很弱。政府越积极推进AI引入，这种联动就越强。

减轻重复规制的机制对现有经营者来说具有实质性意义。根据其他法令已履行相当于第34条第1款各项措施的经营者，按照施行令附件1规定的标准，被视为履行了《人工智能基本法》的责任（第34条第3款）。运营贷款审查AI的金融公司如果遵守了《信用信息法》第35条之2的解释义务并具备解释程序，就被视为履行了解释方案义务。在同一领域，如果全面履行了《金融消费者保护法》第10条的责任，就被视为履行了使用者保护方案。数字医疗器械制造商如果根据《数字医疗产品法》具备了质量管理体系，并获得了软件符合质量管理标准的判定，就被认为履行了风险管理方案，如果还满足了准用的《医疗器械法》规定的质量管理体系要求，就被认为履行了文件制作保存义务。如果是履行了《个人信息保护法》规定的措施，在个人信息处理和保护方面，就被视为履行了相应的责任。

对使用经营者的负担也予以减轻。如果开发经营者履行了针对高影响人工智能的责任，且使用经营者没有显著变更其本来的目的和用途，使用经营者就被视为已履行了风险管理方案、解释方案、使用者保护这三项。人的监督体系的运营与文件保存义务则必须由使用经营者自行履行。开发经营者和使用经营者之间的这种分担，与欧盟将义务分配到整个供应链的方法形成了对照。

欧盟对提供者（provider）施加合格评定义务，对部署者（deployer）施加运营与监视义务，对公共机构和信用评级、保险等部分部署者甚至要求进行基本权利影响评估。双方都承担了相当大的义务。韩国将首要责任集中在开发经营者身上，减轻了使用经营者的负担。这种设计旨在降低引入机构（医院、学校、金融公司）的规制负担。另一方面，当使用经营者将AI系统按照自身背景进行重新构建时，也留下了主张“功能未发生显著变更”的余地。这个边界有多么不明确，未来就有多么容易引发纠纷。

2026年1月22日，五项责任开始产生法律效力并适用于经营者。公告和指南一起发布。指导期正在运转。使用经营者可以向开发经营者索取资料的权利、开发经营者的合作努力义务、对依据其他法令履行的认可等机制开始运转。法律创造的结构正在运转，但这个结构是否足够严密，是一个要由实务来判定的问题。AI基本法支援服务台上堆积如山的提问，正在诉说着那种判定尚未结束。

第12章 高影响人工智能影响评估——非法律强制却为事实上的必须

2026年3月，首尔的一家人工智能初创企业收到了教育部下属机构的招标公告。这是构建初中生试卷成绩预测系统的项目。它属于韩国《人工智能基本法》第2条第4款（次）项规定的“幼儿教育、初等教育及中等教育的学生评估”领域，属于高影响人工智能。翻开招标书，在技术能力、安全认证、个人信息处理业绩等项目后面，出现了一行字：“是否履行人工智能影响评估。”这是一个加分项。法律从未将影响评估规定为强制义务。也没有行政罚款。如果在这个项目中拿到0分，就会在与竞争对手的较量中处于劣势。合规团队负责人下载了科学技术信息通信部的人工智能影响评估指南。

1. 影响评估的法律性质

高影响人工智能影响评估并非义务。

在韩国《人工智能基本法》的任何地方，都没有不实施影响评估就要处以行政罚款的条款。该法第4章在一个法条中规定了高影响人工智能经营者的五项责任，即制定风险管理方案（第34条第1款第1项）、制定解释方案（第2项）、保护使用者（第3项）、人的管理与监督（第4项）、制定及保存文件（第5项），并将其作为义务加以规定，但在该法第35条第1款中，将影响评估本身规定为“应当努力”。在法学上，这种条款被称为“努力义务”。意思不是必须达成结果，而是朝着这个方向努力。

在韩国国会审议过程中，该条款原本是另外一种形式。最初的法案中，有一项条款将事前影响评估规定为强制义务。企业方面提出了反对。他们主张，如果在发布人工智能系统之前还要经过影响评估，市场准入就会延迟，从而在全球竞争中落后。于是达成了一项妥协方案。即将影响评估设为自律的努力义务，同时在公共采购中对履行了该义务的经营者给予优先考虑的结构。它选择了激励的“胡萝卜”，而不是强制的“大棒”。

从法律教科书的分类来看，这一条款更接近于纲领性规范。它向国家和经营者指明方向，但对违反行为没有制裁。即使不实施影响评估，也没有行政罚款，更没有行政处分。

那么，为什么还要做呢？

2. 公共采购之门

答案在于第16条第3款和第35条第2款。

2026年1月20日公布的修订法（法律第21311号）新增了促进公共部门引入人工智能的条款。即第16条第3款。国家机关在采购产品或服务、发包劳务时，必须优先考虑人工智能。将于2026年7月21日施行的修订后《施行令》第15条第4款，将优先考虑的对象规定为经科学技术信息通信部长官确认的人工智能产品和服务，而确认的标准和程序则交由公告来规定。

影响评估就属于这一结构。该法第35条第2款规定，国家机关在拟使用应用了高影响人工智能的产品或服务时，应当优先考虑已实施影响评估的产品和服务。虽然法律没有强制要求，但要进入政府采购市场，接受评估会更为有利。

韩国政府每年的公共采购规模超过180万亿韩元。其中，与人工智能相关项目的比重正在迅速增加。随着修订法第16条第3款将“优先考虑人工智能”法制化，面向政府的市场（B2G）成为人工智能经营者不可忽视的收入来源。同时还加入了公务员免责条款。如果没有故意或重大过失，公务员无需承担因引入人工智能而产生的赔偿责任。公务员对引入人工智能犹豫不决的理由又少了一个。买方的门槛一旦降低，在卖方看来，“影响评估”这一资格要求的权重也就随之增加。

未接受影响评估的经营者并非不能参加投标。但是，与接受了评估的经营者相比，就会处于不利地位。“不做也可以，只是在公共市场上会被边缘化而已。”这就是软监管的设计。

3. 影响评估的三个阶段

科学技术信息通信部于2026年1月22日发布的人工智能影响评估指南提出了三个阶段。在五大指南（确保透明度、确保安全性、高影响人工智能判断、经营者责任、人工智能影响评估）中，这是最后一个。

首先是事前准备阶段。确定作为评估对象的人工智能系统的目的、功能和应用领域。判断该系统是否属于高影响人工智能。判断标准是该法第2条第4款的11个领域和《施行令》第25条的两个阶段程序。在第1阶段，确认是否属于相关领域。即是否应用于能源供应、医疗保健、医疗器械、核设施、生物识别、招聘和贷款审查、交通工具、公共服务、学生评估等领域。在第2阶段，审查是否对人的生命、身体安全及基本权利产生重大影响。如果难以判断，可以要求科学技术信息通信部长官予以确认，长官通常应在30日内回覆结果，但可延长一次，延长期限为30日。组建评估团队，并确定听取外部利益相关者（使用者、受影响者、市民团体）意见的方法。

接着是正式评估阶段。核心是编写风险情景。具体描绘当人工智能系统发生故障或得出有偏见的结果时，会发生什么事情。例如，贷款审查人工智能系统化地拒绝特定地区居民的情景；学生评估人工智能预测残疾学生的成绩低于其实际成绩的情景；招聘人工智能给女性应聘者简历的评分低于男性的情景。不能只描述抽象的风险，必须具体叙述谁受到了伤害，以及是如何受到伤害的。

评估项目涵盖了多项标准。公平性（是否存在偏见和歧视）、透明度（能否解释运行原理）、安全性（发生故障时是否有安全装置）、责任性（发生问题时责任归属是否明确）、个人信息保护（数据处理是否合法）、人为控制的可能性（人类能否通过干预改变结果）。在基本权利影响分析中，要考察对平等权、知情权、隐私权、职业选择自由等宪法权利产生的影响。

最后是事后措施阶段。根据评估结果，制定降低风险的措施。制定监控计划，并定期实施再评估。将评估结果和措施明细以文件形式记录并保存。这与该法第34条第1款第5项及《施行令》第27条第2款要求的履行依据文件需保存5年的义

务相衔接。

4. 由谁来评估

由经营者自己评估。

内部自我评估是基本做法。《施行令》第28条第2款规定，可以由经营者直接实施或委托第三方实施，但并没有规定必须接受外部认证机构验证的义务。让外部专家参与是指南中推荐的事项，但这并非义务。经营者编写自己系统的风险情景，自己对评估项目进行评分，自己制定事后措施，自己制作并保存文件。

对于这一结构是否妥当，存在两种截然不同的观点。

支持自律评估的一方认为：比任何人都了解人工智能系统内部运作的主体就是开发的经营者自己。外部认证机构很难在极短的审查期限内准确把握复杂的机器学习模型的偏见。内部评估在实质上能实现更深层次的分析。如果将外部认证规定为义务，成本就会增加，而在韩国国内缺乏足够的人工智能专业认证机构的情况下，就会出现瓶颈。

反对的一方则认为：如果由经营者自己评估自己所构建的系统的风险，就会产生将风险定得较低的动机。如果影响评估结果不好，就会影响业务开展，因此经营者可能会在编写情景时避重就轻，或者在偏见分析上走形式。而在法律中并没有外部机制可以对此进行确认。

试想一家开发了贷款审查人工智能的金融科技公司实施内部影响评估的场景。假设该公司的模型存在对非首都圈居民统计出较低信用分数的模式。如果在内部评估中如实记录这一模式，作为高影响人工智能的风险就会变高，在与公共机构进行交易时，就会背负需追加解释的负担。如果不予以记录，也不会有任何人去确认。这种动机结构上的漏洞并非法律所期望的，而是自律监管所固有的局限性。

5. 欧盟的合格评定：同名不同刃

欧盟《人工智能法》中的合格评定（conformity assessment）在性质上有所不同。

韩国是鼓励。欧盟则是强制。

欧盟《人工智能法》第43条强制要求高风险人工智能系统的提供者（provider）在推向市场之前必须进行合格评定。如果未通过评定，就无法获得CE标志；如果没有CE标志，就无法在欧盟市场上市。大门就此关闭。

合格评定的路径有两种。大多数高风险人工智能系统采用内部控制（internal control）方式。即提供者自行证明其满足欧盟《人工智能法》第9条至第15条的要件。风险管理体系、数据治理、技术文件、记录保存、透明度信息提供、人为监督机制、准确性及稳健性和网络安全。这七项要求必须全部满足。即使是内部控制，也需要编写技术文件、运营质量管理体系，并签发欧盟符合性声明。

对于用于远程生物识别的人工智能系统，制定了更为严格的路径。如果在未完全适用协调标准的情况下，将由独立的第三方认证机构（notified body）直接审查系统。此时仅靠经营者的内部自我评估是不够的，必须引入外部监督机制。

欧盟《人工智能法》第27条单独规定了基本权利影响评估（Fundamental Rights Impact Assessment, FRIA）。这与合格评定是两项不同的义务。公共机构或提供公共服务的私营企业在部署高风险人工智能时，必须在部署前实施基本权利影响评估。这不是提供者（provider）的义务，而是部署者（deployer）的义务。评估需要将受影响人群的范畴、使用期限、具体风险、人为监督措施以及对负面影响的应对方案记录在案。其结果必须通报给国家监督机构。

概括起来就是：欧盟将两项评估分离开来。对提供者要求合格评定，对部署者要求基本权利影响评估。这两项都是强制义务。韩国没有设立类似欧盟的合格评定，而是将单一的影响评估设定为自律的努力义务。

6. 五大决定性差异

韩国的影响评估与欧盟的合格评定虽然名字相似，但在几乎所有方面都截然不同。

法律性质不同。韩国是努力义务。欧盟是市场准入的法定前提条件。在韩国，如果不进行影响评估，会在公共采购中处于不利地位。在欧盟，如果不进行合格评定，则根本无法进入市场。

评估主体不同。韩国是由经营者直接实施或委托第三方实施的自律评估，不强制要求外部验证。欧盟虽大部分采用内部评估（内部控制），但在远程生物识别等敏感领域，会引入第三方认证机构。

结果的形式不同。韩国的影响评估结果是经营者保存的内部文件。如果在欧盟通过合格评定，则会在产品上贴上称为“CE标志”的物理标识，并签发欧盟符合性声明。这是市场和消费者可以用肉眼确认的信号。

违规制裁不同。在韩国，不进行影响评估不会被处以行政罚款。在欧盟，如果未经合格评定就推出高风险人工智能，最高将被处以1500万欧元或全球营业额3%的罚款。包括违规粘贴CE标志在内的违反提供者义务的行为，也属于同一处罚区间。如果违反了被禁止的人工智能实践，罚款将高达3500万欧元或营业额的7%。在韩国《人工智能基本法》整体设定的最高3000万韩元行政罚款，与欧盟的3500万欧元之间，存在四个数量级的差距。

适用时间不同。韩国的人工智能影响评估指南从2026年1月22日开始适用。欧盟关于高风险人工智能合格评定的义务，一旦《数字综合修订案》最终确定，预计将推迟至2027年12月2日。虽然韩国先采取了行动，但一旦欧盟真正开始实施，其监管力度将具有压倒性的分量。

7. 韩国选择自律监管的理由

这是有意的选择。

韩国《人工智能基本法》的正式名称是《关于人工智能发展与奠定信任基础等基本法》。“发展”排在“信任”前面。这部法律是振兴法与监管法的混合体，而重心放在了振兴这一侧。

在韩国国会审议当时，产业界的逻辑是这样的：韩国人工智能产业在全球市场上属于后来者。如果对那些必须与中美大型人工智能企业竞争的韩国初创企业和中小企业强加事前合格评定，就会减缓其进入市场的速度，并增加成本。就连欧盟也通过《数字综合修订案》，提出了将高风险人工智能义务推迟16个月至2027年12月的方案，并提出了削减中小企业监管成本35%的目标。在欧盟都在放慢自己监管步伐的情况下，韩国有什么理由率先制定严厉的监管措施呢？这就是他们的逻辑。

科学技术信息通信部也确定了同样的方向。由80多名民间专家组建下级法令整顿小组，确立了由经营者自律确保安全性为基本原则。5项指南和2项公告（最前沿人工智能确保安全性公告、高影响人工智能经营者责任公告）均遵循了这一原则。指南不具法律约束力。它作为实务工作指针，为经营者展示履行法定法律义务的方法。

行政罚款的额度也是出于同样的考量。最高3000万韩元。这是基于韩国国内人工智能产业仍处于成长初期的判断。在法案审议的整个过程中，一直有人担忧高额罚款会阻碍创新。

激励机制的设计也源于这一理念。也就是说，与其通过法律进行惩罚，不如通过市场去吸引。如果实施影响评估，在公共采购中就会更加有利。可以在人工智能基本法援助服务台进行匿名咨询。在宣导期内（至少一年，到2027年1月），优先采取行政指导和纠正建议，而不是行政罚款。这是一种铺设了多个“胡萝卜”，暂时放下“大棒”的结构。

8. 这一设计的漏洞

批评之声来自多个方面。

韩国法制研究院研究委员郑元俊在2025年9月的报告（《人工智能基本法的合理监管及发展方案研究》）中指出，“高影响人工智能”这一概念缺乏严密性。在法律条文中，“影响”与“风险”混合使用，“安全”与“基本权利”用“及（and）”连接，这就造成了解释上的混乱。即使经营者想要进行影响评估，如果什么是“重大影响”的标准模糊不清，评估本身就会流于形式。

同一份报告还指出，由于监管对象过于宽泛，导致政府机关也被纳入了“人工智能经营者”范畴。这是因为该法没有像欧盟《人工智能法》那样按照价值链（provider, distributor, deployer, user）对守法主体进行细分，而是将其笼统地划分为开发经营者和使用经营者两大类。影响评估的主体和范围就建立在这种模糊性之上。

韩国国会立法调查处研究官郑俊华在法律实施日即2026年1月22日发布的报告（《为实现国家人工智能战略的行政课题》）中指出，具体化高影响人工智能的概念和要件、根据经营者类型进行义务差别化管理是剩下的课题。

韩国法制研究院研究委员李有奉在2024年10月的《用于人工智能影响评估的立法研究》中，对欧盟与美国的影响评估制度进行了比较分析，并建议为确保韩国型影响评估的实效性，必须具体设计评估结果的公开范围和外部验证体系。

内部自我评估结构本身就存在漏洞。即使有经营者故意将风险情景写得很低，或者在偏见分析上敷衍了事，法律中也没有能够对此进行确认的机制。这与欧盟对远程生物识别人工智能要求第三方认证机构评估形成了鲜明对比。

最高3000万韩元的行政罚款也存在威慑力不足的问题。对于每年创造成百上千亿韩元营业额的全球人工智能企业来说，区区3000万韩元根本不会被视为一笔经营成本。欧盟之所以将营业额的7%设定为罚款上限，是因为他们认为，只有当罚款数额重到足以影响商业决策时，监管机制才能真正发挥作用。

9. 软监管是否有效？

这个问题的答案尚未浮出水面。

软监管要发挥作用是有条件的。即市场参与者必须理解法律的宗旨并自愿遵守，或者当不履行义务时在市场上遭受的不利益必须足够大。公共采购市场必须足够庞大，并且影响评估的履行与否必须在采购评估中占据实质性的比重。

支持的一方这样说道：韩国政府的人工智能政策正在全方位地推进。国家人工智能战略委员会已经敲定了包含99项执行课题和326项政策建议的行动计划。48个中央行政机关已全员指定副部级人工智能最高责任官（CAIO），并每月举行例会。韩国智能信息社会振兴院（NIA）内设立了人工智能政策中心，由59人负责政策开发和影响分析。一旦这个体系运转起来，在公共机构的人工智能采购中，是否履行影响评估自然就会成为一个重要的评估标准。即使法律不作强制要求，市场也会强制执行。

反对的一方则这样说道：依赖市场自律的监管，在市场失灵的领域必然会失败。对于对B2G市场毫无兴趣的经营者、仅运营面向消费者（B2C）服务的经营者而言，公共采购激励机制起不到任何作用。对于自行运营招聘人工智能的大企业、仅在内部使用贷款审查人工智能的金融机构来说，公共采购加分项也完全不构成任何影响。法律中没有任何动力能促使这些经营者自发实施影响评估。

人工智能基本法援助服务台的业绩数据可供参考。在实施仅10周的时间里，就收到了552件咨询。近一半的咨询集中在透明度义务方面，其次是高影响人工智能确认问题。这是企业经营者有意愿遵守法律的信号。然而，咨询案件的数量与实际实施影响评估的数量是两码事。询问与实际行动之间，还存在着距离。

审视这三个国家和地区的路径，各自的选择十分鲜明。欧盟选择了硬监管。如果不通过合格评定，就无法进入市场，一旦违规，将按全球营业额百分比开出罚单。中国选择了事前注册（备案）。只要推出具有舆论属性或社会动员能力的生成式人工智能服务，必须在10个工作日内对算法进行备案。截至2025年11月，已有611项服务完成了备案。如果不进行备案，将面临责令整改，甚至可能导致

服务暂停。而韩国则选择了激励机制。即使不做也没有惩罚，如果做了，在公共采购市场上就会处于有利地位。

究竟哪种路径才是正确的，在2026年7月还无法下定论。欧盟关于高风险人工智能合格评定的义务要到2027年12月才正式实施。韩国将评估与公共采购相挂钩的结构，将于2026年7月21日起生效。实证比较恐怕要等到几年之后才能进行。

不过，有一个趋势已经清晰可见。既然48家机构的CAIO每个月都会召开会议，99项课题正在推进，而且第16条第3款已经以法律形式将“优先考虑人工智能”板上钉钉，那么未经影响评估的人工智能系统在公共招标中被选中的可能性，势必会逐年递减。

首尔的那家初创企业开始实施影响评估。这不是因为法律规定，而是为了在竞标中胜出。它身旁的竞争对手也在进行同样的算计。这就是没有任何强制的监管开始发挥作用的瞬间。这一作用能否延伸到B2G以外的市场，能否防止内部评估流于形式，以及3000万韩元的行政罚款能否形成足够的威慑力？法律所留下的这些问题，正等待着市场给出答案。

第13章. 确认是否属于高影响的申请程序

2026年3月，在首尔江南区一家金融科技公司的会议室里，法务组长展示了一张幻灯片。“我们的贷款审查AI，确实属于高影响。”开发组长反问道：“因为是贷款审查领域，所以自动就是高影响吧？为什么还要另外申请确认呢？”法务组长摇了摇头。“属于该领域并不意味着直接就是高影响。有两阶段的判断。而且如果不申请确认，以后我们会很麻烦。”会议室安静了下来。在场的人中，有一半是第一次知道必须向科学技术信息通信部提交确认申请书。

1. 什么是确认申请

《人工智能基本法》规定了运营高影响人工智能的经营者需要承担制定风险管理方案、说明义务、确保人工干预体系、文件保存5年等责任。是否应该承担这些责任，取决于自己的服务是否属于高影响。确认申请是向科学技术信息通信部正式询问是否属于高影响的程序。

经营者是主体。这并不是政府主动找上门来通知“你的人工智能是高影响”的结构。该法第33条第1款规定，经营者有义务事先审查自己的人工智能是否属于高影响。如果审查后仍不确定，可以向科学技术信息通信部长申请确认。在这一点上，韩国的制度以经营者的事先审查为出发点，并在判断困难时设立了由政府给出答案的辅助机制。

确认申请的对象包括开发经营者和使用经营者。对自己开发的AI模型是否在经过高影响领域使用感到好奇的开发经营者，以及接收外部模型并将其应用于贷款审查或招聘评估的使用经营者，都可以走相同的程序。

2. 程序流程：30日，以及30日

确认申请被受理后，科学技术信息通信部原则上会在30日内通报结果。考虑到产品或服务的复杂性和重要性等，可以延长一次30日，此时必须以书面形式告知延长事由和延长期限。最长处理期限为60日。

30日这个期限在施行令制定过程中被多次讨论。产业界曾主张应缩短至15日。论据很明确。在AI服务发布周期不断加快的环境下，30日几乎等于一个时代。如果在竞争对手已经推出服务后才收到确认结果，就会错失市场机会。对于初创企业来说，30日的空档期可能会与吸引投资的日程发生冲突。

也有反对的论据。高影响的判断必须同时考量技术特性和社会影响力。医疗领域的AI和能源供应领域的AI需要审查的安全标准完全不同。有些情况还需要专家咨询。反驳意见认为，很难在15日内对所有领域进行精确判断。

结果是维持了原则上的30日。不过，法律并没有规定期限的下限，因此，如果争议较少的明确事项，可能不到30日，仅在10日内就会得到答复。

3. 两阶段判断的结构

收到确认申请的科学技术信息通信部适用的判断标准是两阶段。

第一阶段是领域确认。查看该AI系统是否在《人工智能基本法》第2条第4项列举的11个领域之一中使用。这些领域包括：能源供应，饮用水生产过程，医疗保健，医疗器械及数字医疗器械，核材料与核设施的安全管理，用于犯罪调查或逮捕业务的生物识别信息的分析与使用，招聘或贷款审查等对个人的权利和义务产生重大影响的判断，交通工具与交通设施的主要运行管理，提供公共服务所需的资格确认·决定和费用征收等对国民产生影响的国家机关等的决策，幼儿教育·初等教育·中等教育中的学生评估，以及可以通过总统令追加指定的领域。

第二阶段是重大性判断。仅仅在相关领域中使用，不足以成为高影响。必须存在“对人的生命、身体安全及基本权利产生重大影响或引发危险的担忧”。即使在贷款审查领域使用AI，如果该AI并不做出最终决定，而仅仅是作为整理参考资料的辅助工具，则可能不符合重大性标准。相反，如果AI事实上决定贷款批准与否，且负责人几乎不推翻该结果的结构，则很有可能被认定为具有重大性。

高影响人工智能判断指南为经营者提供了在进行这两阶段判断时可以参考的具体标准。指南本身虽然没有法律约束力，但科学技术信息通信部在审查确认申请时会以此指南为基准进行判断，因此它是事实上的参考规范。

4. 自主核对表与低影响申报制度讨论

在施行令立法预告过程中，出现了一种替代方案的意见。即对于不属于高影响的AI，即低影响AI，引入简易申报制度的讨论。

概念是这样的：由科学技术信息通信部提供自主核对表。经营者逐项对照核对表，如果自主判断不属于11个领域中的任何一个，或者即使属于也不满足重大性标准，则将该结果向科学技术信息通信部申报。这是一种无需经过30日的确认程序，仅通过申报即可获得“非高影响”确认的方式。

提出这一讨论的背景存在现实担忧。在韩国，提供AI服务的经营者数量正急剧增加。如果所有经营者都向科学技术信息通信部提交确认申请，行政力量将难以承受。让明显属于低影响的事项也走30日的程序，对经营者和政府来说都是低效的。

截至2026年7月，自主核对表已部分反映在指南中，但正式的低影响申报制度尚未制度化。在经营者自行判断不属于高影响的情况下，将判断依据以文件形式留存是实务上安全的应对措施。因为以后进行事实调查时，需要证明为什么认为它不是高影响。

5. 确认结果的法律效力

如果科学技术信息通信部回复“属于高影响”，该经营者必须承担该法第34条第1款各项义务。必须制定并运营风险管理方案。必须制定并实施用于说明AI得出的最终结果、得出该结果所用的主要标准、训练数据概要的方案。必须制定并运营使用者保护方案。必须确保由人进行管理和监督的体系。必须编写能够确认措施内容的文件并保存5年。同时，根据施行令第27条第1款，应在办事处、营业场所或互联网主页等处公布风险管理方案的主要内容、说明方案的主要内容、使

用者保护方案、负责管理·监督人员的姓名和联系方式，但属于商业秘密的事项可以排除。

“不属于高影响”的回复具有什么效力呢？这意味着以当时的该服务形态为基准，确认无需承担高影响人工智能的相关义务。但这并非豁免。如果服务发生改变，判断也可能改变。如果曾经作为贷款审查辅助工具的AI经过版本升级，变成了事实上的自动审查系统，那么经营者必须以变更后的服务为基准，重新审查是否属于高影响，并在必要时重新申请确认才是安全的。

当经营者不同意属于高影响的判断时，可以先申请重新确认。根据施行令第25条第4款和第5款的规定，自收到回复之日起10日内提交重新确认申请书，科学技术信息通信部长将听取专门委员会的意见，并在收到重新确认申请之日起30日内回复结果。如果对重新确认结果仍不同意，是否可以提起行政争议便成为焦点。法律专家的普遍解释是，如果确认回复属于行政处分，则可以通过行政复议或行政诉讼进行争议。

6. 被确认为高影响后的后续程序

如果确认结果为“高影响”，经营者必须依次采取各项措施。

制定风险管理方案是第一步。科学技术信息通信部的指南将风险管理的程序分为五个阶段：制定政策、识别风险、分析与评估、处理、改善。这与ISO 31000（风险管理国际标准）的结构相似。这是一个了解自己的AI系统包含哪些风险，分析各风险的严重程度和发生可能性，采取措施降低风险，并监控其结果以进行改善的循环结构。

建议进行影响评估。对于高影响AI经营者，基本权利影响评估并非强制义务，而是“自主努力义务”。虽然不受法律强制，但进行了影响评估的经营者在公共采购中会获得优先考虑。对于计划进入政府采购市场的经营者来说，影响评估实际上是必不可少的。

透明度义务同时并行。必须告知使用者正在运营高影响AI的事实，并且为了使受到影响的人能够获得关于AI做出判断的主要标准和原理的说明，必须制定并实施说明方案。

文件化义务在时间上具有长尾效应。履行依据文件必须保存5年。由于施行令第27条第2款未另行规定保存期限的起算点，因此即使服务终止，在自履行各义务之时起满5年之前，妥善保管文件才是安全的。

7. 欧盟的合格评定：自我声明和第三方评定的二元结构

欧盟《AI Act》的高风险AI体系与韩国有一个根本不同的地方。欧盟并不要求经营者向政府询问“请确认我的人工智能是否是高风险”。经营者必须自行判断，并通过合格评定（conformity assessment）后才能投放市场。

合格评定有两条路径。一条是自我声明（self-declaration）。经营者自行声明满足欧盟《AI Act》的要求，编写技术文件，并加贴CE标志。附件III中列出的大多数高风险AI系统都遵循此路径。

另一条是第三方评定（third-party conformity assessment）。由独立的合格评定机构（Notified Body）检验相关AI系统。要求采用此路径的范围很窄。附件III的生物识别系统中未完全适用协调标准或通用标准的情况，以及像机械或医疗器械一样与现有产品安全法令相关联的类型属于这一类。核心基础设施或执法等其余的高风险领域原则上采用内部控制方式，即自我声明路径。

韩国不属于这两条路径中的任何一条。韩国的高影响确认是经营者向政府提问的结构，而欧盟的合格评定则是经营者直接证明或由第三方验证的结构。韩国看重“影响”而欧盟看重“风险”的哲学差异，在这里体现为程序上的差异。

制裁规模的差距支撑了这种结构性差异。在欧盟，如果没有进行合格评定就将高风险AI投放市场，最高可被处以1500万欧元或全球营业额3%中较高数额的罚款。韩国则不同。根据该法第43条，行政罚款仅在未进行事先告知、未指定韩国

境内代理人、不履行停止命令或纠正命令时才被征收，上限为3000万韩元。违反高影响经营者义务本身并不会直接招致行政罚款，而是只有在接到纠正命令后仍不履行时，才会导致行政罚款。这是在进入市场前必须证明的体系与进入市场后进行确认的体系之间的鸿沟。

8. 支持热线的17%

人工智能基本法支持热线自2026年1月22日开通以来，一直在接收经营者的提问。查看咨询内容就能看出法律的现实。

开通后前十天收到的94件在线咨询中，最多的类型是关于透明度义务的，占56.4%。“应该以什么形式告知正在使用AI的事实呢？”“生成物标识应该贴在哪里，怎么贴？”就是这类问题。其次是高影响AI确认，占17%，第三是关于经营者定义的咨询，占10.6%。比如“我们是开发经营者，还是使用经营者”这类问题。

在实施到第10周时累计的552件咨询中，结构也是一样的。透明度相关的咨询约占51%，依然占据一半，而高影响确认约占19.6%，约108件，比重反而增大了。

从17%到19.6%的这个数字说明了一个问题。这意味着十个经营者中几乎有两个无法判断自己的服务是否属于高影响。这些经营者大概都读过该法第2条第4项的11个领域列表。读了列表依然无法做出判断的原因在于第2阶段的重大性判断。虽然属于相关领域，但很难自行确定自己的AI是否对基本权利产生了“重大影响”。

支持热线在工作日上午9点至下午6点运营。需要深入法律审查的事项最长需要14日，如果审查延迟，最多需要21日。也可以不透露实名或公司名称，用假名进行咨询。这是一个不泄露商业秘密也能确认自身服务法律地位的渠道。

9. 推迟确认申请的经营者的风险

确认申请并非义务。但是，由于该法第33条第1款将事先审查是否属于高影响本身规定为经营者的义务，因此甚至连审查都不做就置之不理并不是法律预期的状态。经营者在自行审查后判断“我们的服务不是高影响”，并决定不申请确认，这本身并不违法。问题在于该判断出错的时候。

经营者自行判断不属于高影响后，如果日后在科学技术信息通信部的事实调查中被查明属于高影响，会发生什么呢？该经营者将处于未履行赋予高影响AI经营者的所有义务的状态。没有风险管理方案，没有说明体系，文件也没有保存。在实施初期，很可能会优先进行行政指导和纠正劝告，但如果下达了根据该法第40条的纠正命令却不履行，可能会面临3000万韩元以下的行政罚款。

更大的风险在于商业成本，而非法律制裁。溯及性地具备高影响AI经营者义务的成本，比一开始就具备要高出好几倍。必须重新制定风险管理方案，对过去期间的记录进行文件化可能变得不可能，甚至可能需要改变服务设计。在运营服务的同时建立合规体系，无异于在行驶的车上换轮胎。

在公共采购市场中也会处于不利地位。该法第35条第2款规定国家机关等应优先考虑进行了影响评估的产品和服务，因此如果连是否属于高影响都未能理清，从而未能触及影响评估的经营者，在竞争中可能会落后。

还有一个风险。当使用者或受影响者要求民事损害赔偿时，经营者未进行高影响确认申请的事实可能会被作为存在过失的依据。其逻辑是：“如果是一个理性的经营者，应该会去确认自己的服务是否属于高影响。”法院将如何判断这一点目前尚无判例，但未申请确认对经营者而言很难成为有利的情节。

10. 高影响判断指南的实务适用

该指南于2026年1月22日作为5大指南之一发布。它没有法律约束力。既不是行政规则，也不是公告。然而，如果科学技术信息通信部在审查确认申请、进行事实调查时，均依照指南进行判断，经营者就无法无视这一指南。

指南给出的判断流程具体化了施行令的两阶段结构。经营者首先明确自己AI系统的使用目的和应用领域。确认是否属于11个领域之一。如果属于相关领域，就进入下一阶段。考察该AI系统在相关领域扮演什么角色、对最终决定参与到什么程度、人类的监督在什么水平上发挥作用。

判断依据必须以文件留存。指南建议记录“为什么判断为高影响”或“为什么判断为非高影响”的依据。这份文件日后可能成为确认申请的申辩资料，也可能成为在事实调查中证明经营者的自主判断具有合理性的证据。

实务中判断棘手的地方在于，AI虽然作为辅助工具使用，但事实上深度参与了决策的情况。假设在招聘过程中有一种对简历进行打分并排名的AI。最终的招聘决定由人事主管做出。但是，如果人事主管几乎从未推翻过AI的排名，那么这个AI事实上就是在做决定。指南引导应以这种实质性的功能为基准进行判断。

11. 韩国与欧盟程序的结构性差异

将两个体系并列放置，可以看到监管哲学差异转化为程序的过程。

韩国的经营者向科学技术信息通信部询问：“这个是高影响吗？”政府予以回答。欧盟的经营者则是自行判断，经过合格评定后，向市场声明：“这个是安全的。”韩国是政府作为判定者，而欧盟是经营者作为证明者。

在欧盟的自我声明路径中，经营者编写技术文件（technical documentation），建立质量管理体系（quality management system），并签署欧盟合格声明（EU Declaration of Conformity）。只有备齐这些文件，才能贴上CE标志并投放市场。市场监督机构（Market Surveillance Authority）会事后检查这些文件。

在欧盟的第三方评定路径中，由认证机构（Notified Body）直接检验AI系统。像生物识别系统这样基本权利侵犯风险较高的类型就会走这条路径。这是基于仅靠经营者的自我声明是不够的判断。

韩国不存在这两条路径中的任何一条。合格评定这个概念本身并没有被引入《人工智能基本法》。韩国的确认程序仅仅是判别“是否属于高影响”，而并没有进一步证明“是否安全”。取而代之的是，韩国试图通过包括制定风险管理方案在内的该法第34条第1款的法律义务，以及作为自主努力义务的影响评估，来确保安全性。

现在断言哪一种更好还为时过早。欧盟方式在进入市场前筛选出安全性问题，但经营者负担较大，可能会减缓创新速度。通过综合修正案，独立型高风险AI的义务期限被推迟到2027年12月，而内置于产品中的类型则被推迟到2028年8月，这就是该负担的证据。韩国方式虽然不阻碍进入市场，但当事后爆发问题时，如果经营者的自主判断不充分，制裁手段就会偏弱。3000万韩元的行政罚款上限能否对大企业发挥威慑力的问题仍未解决。

12. 无法做出判断时

是否属于高影响的确认申请，其核心在于“无法做出判断的领域”。对于核设施安全管理AI或用于犯罪调查的生物识别AI这样明显的情况，可能根本不需要申请确认。经营者自行判断为高影响，并履行义务即可。

问题在于灰色地带。在教育领域，追踪学生学习进度的AI是高影响吗？该法第2条第4项J目列举了“幼儿教育、初等教育及中等教育中的学生评估”。追踪学习进度是评估还是监控呢？在医疗保健领域推荐饮食食谱的AI又如何呢？这可能属于医疗保健“提供与使用体系的构建和运营”，但是否能认为推荐食谱会对人的生命或身体安全产生重大影响呢？

这种处于边界的事项占据了支持热线约五分之一的咨询量。在法律实施6个月后的2026年7月，随着经过确认申请程序的案例及其结果的积累，判断标准将会被具体化。在初期，实际确认结果的先例比指南的文字对经营者来说是更有用的参考资料。

确认申请不是成本，而是保险。拿到了科学技术信息通信部正式回复的经营者，在日后需要证明判断正当性的情况下就有了依据。推迟确认申请的经营者，就只能祈祷自己的判断是对的。

第五部 经营者面临的相邻法律风险

第14章. 训练数据与著作权：合理使用条文、韩国首起诉讼、海外判例的警告

2023年12月27日，《纽约时报》向美国纽约南区联邦地区法院递交了诉状。被告栏上写着OpenAI和微软。在诉状附带的近100页证据中，引人注目的是ChatGPT几乎逐字逐句再现《纽约时报》新闻报道的屏幕截图。有关出租车牌照危机的调查报道在几个句子的单位上完全一致。辩护律师将损害赔偿金额写为“数十亿美元”。这是用于人工智能训练的数据的著作权问题到达法庭的瞬间。

1. 所谓训练数据的一行字

韩国《人工智能基本法》第2条第12项对“训练数据”进行了定义。“指用于人工智能的开发、使用等的数据。”只有一行字。在十二个定义中字数最少，留白领域最广。

该定义并未限定数据的种类。文本、图像、语音、视频、代码、医疗记录、法律文书、新闻报道、小说。只要是投入到制造或使用人工智能中的数据，全都是训练数据。属于谁的、如何收集的、是否获得了许可。该定义都不予过问。

韩国《人工智能基本法》第15条第4款规定了构建训练数据综合提供系统的依据，《施行令》第13条规定了该系统的功能和管理方式。应当具备训练数据的综合检索、系统分类、出处提供功能。这是以公共数据为中心提供用于人工智能训练的数据的基础设施。然而，在将非公共数据的民间著作物用于训练时需要何种条件，韩国《人工智能基本法》并未提及。在韩国《人工智能基本法》的“训练数据”定义与《著作权法》的保护体系之间不存在连接环节。在这片空白空间中，诉讼开始了。

2. 所谓合理使用的四杆秤

《著作权法》第35条之5规定了“著作物的合理使用”。这是在2011年韩国《著作权法》修订时引入美国《著作权法》第107条合理使用（fair use）法理的条

款。

第1款规定：“在不与著作物的一般使用方法相冲突，且未不当损害著作权人正当利益的情况下，可以使用著作物。”第2款规定在判断此时要考虑四点。使用的目的及性质。著作物的种类及用途。被使用的部分在整个著作物中所占的比重及其重要性。著作物的使用对该著作物的当前市场或价值、以及潜在市场或价值的影响。

将这四项标准代入人工智能训练中，答案并不明确。

使用的目的及性质。人工智能模型训练具有商业性。OpenAI、NAVER、Kakao全都为了商业服务而收集数据。营利性使用不利于合理使用的认定。相反，也有反驳观点认为，在训练过程中并非享受著作物的表现形式本身，而是提取统计模式。如果不是“阅读”著作物而是“分析”著作物，则主张该使用的性质具有转换性（transformative）。

著作物的种类及用途。新闻报道、小说、学术论文、照片各自的保护程度不同。接近事实报道的新闻保护范围较窄，小说或诗歌的保护范围较宽。由于人工智能训练不分体裁进行大量收集，该标准无法一律适用。

被使用的比重。人工智能模型复制整个著作物。当把一篇新闻报道放入训练时，放入的是全文而非部分。整体复制不利于合理使用的认定。

市场影响。第四个标准是刀刃。当人工智能训练了新闻报道后，生成类似内容提供给使用者，使用者访问原新闻网站的理由就会减少。其逻辑是著作物的潜在市场被蚕食。人工智能经营者方面反驳称，训练数据和生成物之间不存在直接的替代关系。

韩国法院尚未就人工智能训练数据是否构成合理使用做出过判决。这四杆秤的刻度将定在何处，要等判决出炉时才能确定。

有一点需要注意。韩国《著作权法》中没有像日本《著作权法》第30条之4（信息解析目的使用）那样明示允许人工智能训练或文本和数据挖掘（TDM）的条款。仅仅只有合理使用的一般条款（第35条之5）。一般条款需要法院在各个案件中对四项标准进行权衡，因此无法给经营者提供可预测性。

3. 首尔法院的战线

在韩国，人工智能训练数据著作权纠纷的第一炮是由广播公司打响的。

2025年1月，KBS、MBC、SBS等三大地面广播公司以NAVER为被告向首尔中央地方法院提起了著作权侵权诉讼。广播公司方面主张，NAVER在训练HyperCLOVA X等韩语大型语言模型时，未经许可大量使用了广播新闻内容。

2026年2月，上述三家广播公司又以OpenAI为被告向首尔中央地方法院提起诉讼。请求原因在于ChatGPT将广播新闻内容作为训练数据使用而未获得许可。战线从韩国国内门户网站扩大到了海外人工智能企业。

该纠纷向《著作权法》抛出的问题是明确的。第35条之5的合理使用能否适用于商业性人工智能训练。在将大量新闻内容用于训练之后，若人工智能生成了类似的新闻摘要或答复，是否会替代原著作物的市场。

门户网站经营者方面可以主张人工智能训练并非提取著作物的表现形式，而是提取事实和模式的行为，因此属于合理使用。广播公司方面将对此进行反驳，认为本公司内容被用作人工智能服务的核心原料，而该人工智能服务替代了新闻消费，因此发生了市场蚕食。

这起诉讼的结果将决定韩国人工智能产业的数据获取结构。如果法院认可合理使用，就可以相对自由地将互联网上公开的内容用于训练。如果否定合理使用，就必须对所有训练数据获得个别许可。该费用对大企业也是一种负担，对初创企业来说更是高墙。

4. 纽约的两场战争

在美国，有两起诉讼处于人工智能训练数据著作权争论的中心。

《纽约时报》诉OpenAI案。2023年12月27日向纽约南区联邦地区法院提起。

《纽约时报》主张OpenAI和微软未经许可将数百万篇新闻报道用于GPT模型的训练。诉状的核心证据是ChatGPT几乎原样输出《纽约时报》新闻报道的案例。人工智能甚至再现了付费订阅墙后的报道。《纽约时报》要求赔偿数十亿美元的损失，并发布永久禁令要求在GPT模型中删除本公司数据。

OpenAI提出了合理使用抗辩。其逻辑是，即使复制新闻全文，在训练过程中学习的并非著作物的表现形式而是语言的统计结构，因此属于转换性使用。并反驳称新闻再现案例是模型的故障（regurgitation），而非设计出的功能。

在诉讼进行期间，OpenAI开始与其他新闻媒体签订许可协议。美联社、阿克塞尔·斯普林格（德国）、《世界报》（法国）等签订了内容使用协议。发起诉讼的新闻媒体与签订协议的新闻媒体并存。这意味着对于人工智能经营者而言，训练数据许可可以成为不可避免的成本。

美国作家协会（Authors Guild）诉OpenAI案。2023年9月，美国作家协会与约翰·格里沙姆、朱迪·皮考特、乔治·R.R.·马丁等著名作家向纽约南区联邦地区法院提起了集体诉讼。请求原因在于GPT模型未经许可将数千本书籍用于训练。主张著作被全文复制投入到人工智能中，这属于侵犯复制权。

该诉讼中的争议焦点是转换性使用（transformative use）法理。美国联邦最高法院在2023年Andy Warhol Foundation v. Goldsmith案中缩小了转换性使用的范围。即在将原著作物用于商业目的的同时，很难主张其具有转换性。虽然沃霍尔判决并非判断人工智能训练本身的案件，但可以被援引为缩小针对商业目的的大量复制提出转换性使用抗辩的论据。

然而，2025年加利福尼亚北区联邦地区法院做出的两项判决却展示了不同的方向。在Bartz v. Anthropic案中，法院认定用合法购买的书籍训练语言模型本身属

于高度转换性的合理使用。不过，下载并保管盗版图书的行为被视为单独的侵权行为，Anthropic在2025年9月与作家们达成了规模约15亿美元的和解。在Kadrey v. Meta案中，法院在与相关原告的关系中也接受了Meta的合理使用抗辩。其趋势是，承认训练阶段的转换性，但对于数据是如何获取的以及对市场产生何种影响则另行考量。

截至2026年7月，纽约的两起诉讼均未达到终审判决。然而，诉讼本身已经在改变市场。全球性人工智能企业开始将训练数据许可费用反映到业务计划中，也有企业采取了转向合成数据（synthetic data）的战略。

5. 东京敞开的大门

日本在人工智能训练数据问题上，拥有比任何国家都宽容的法律框架。

日本《著作权法》第30条之4规定，在“不以自己享受或让第三人享受著作物中表达的思想或感情为目的的情况下”，可以使用著作物。这是在2018年修订《著作权法》时引入的。信息解析、机器学习、文本和数据挖掘一直被作为该条款的典型适用案例被提及。

逻辑是这样的。当人工智能训练小说时，并非在享受小说的表现形式。而是提取语言的统计模式。由于享受并非其目的，因此属于著作权限制事由。根据这种解释，为人工智能训练进行的大量复制原则上是允许的。

2024年，日本文化厅下属文化审议会著作权分会重新审视了人工智能与著作权之间的关系。争议焦点有两点。在生成与人工智能所训练的著作物类似的结果物的情况下，是否也适用第30条之4。大规模商业性人工智能训练是否不属于“不当损害”著作权人利益的情况（同条但书）。

日本文化厅公开了《关于人工智能与著作权的考量事项》报告书。提出了双重结构：通过第30条之4允许训练阶段的复制，但在生成阶段若输出与原著作物相似的表现形式，则可能构成著作权侵权。这是一种敞开训练、过滤生成的路径。

日本给韩国留下了教训。日本虽然设立了明示条款，但围绕其范围再次陷入了争论。韩国则根本没有明示条款。仅凭借合理使用一般条款来处理人工智能训练，其法律不确定性比日本更大。

6. 布鲁塞尔的公开原则

欧盟选择了在训练数据问题上要求透明度的方向。

《EU AI Act》第53条第1款对通用人工智能（GPAI）模型提供者赋予了制定并履行遵守欧盟著作权法政策的义务。具体而言，应当识别并遵守根据《数字化单一市场版权指令》（DSM Directive, 2019/790）第4条第3款规定的著作权人权利保留。

《DSM指令》第4条对文本和数据挖掘（TDM）做出了例外规定。可以为了TDM目的复制并提取能够合法访问的著作物。不过，在著作权人以机器可读的方式拒绝使用（opt-out）的情况下，不适用该例外。robots.txt文件是其代表性的拒绝手段。

robots.txt是放在网站根目录下的文本文件。从1994年起，一直被用于指示搜索引擎爬虫不要收集某些页面。欧盟将这项有30年历史的技术标准提升为了著作权保护的法律工具。

2025年8月1日，欧盟批准的GPAI《行为准则》（Code of Practice）使这一结构具体化。《行为准则》由透明度、著作权、安全与安保三大支柱构成。GPAI模型提供者在收集训练数据的过程中，必须尊重robots.txt等权利保留标识，并应当将模型相关文档保存10年。公开关于用于训练的内容的充分详细的摘要，是先于《行为准则》由《AI Act》第53条第1款直接赋予的义务。

与韩国相比，差异是鲜明的。欧盟在著作权法中明示TDM例外的同时，也保障了opt-out权利。人工智能经营者可以在能够合法访问的内容中，将权利人未通过robots.txt等予以拒绝的内容用于训练，被拒绝的内容则不能使用。规则相对

明确。韩国既没有TDM例外，也没有opt-out制度。这种结构给经营者留下了法律不确定性，给著作权人留下了保护空白。

7. 空白的空格

韩国国会立法调查处的立法调查官白智妍于2025年5月7日发布了《生成式人工智能训练数据公开相关讨论及立法课题》（《热点与论点》第2355期）。该报告书指出，韩国没有要求生成式人工智能经营者公开训练数据出处与构成的法律依据。

欧盟规定GPAI模型提供者有公开训练数据摘要的义务。在美国，通过诉讼事实上正在施加公开的压力。日本虽然设立了明示的例外条款，但正在重新讨论其局限性。韩国在《人工智能基本法》中定义了训练数据，但并未规定有关该数据收集方法或出处公开的内容。《著作权法》中也没有关于为人工智能训练进行复制的明示条款。

剩余的立法课题是鲜明的。是否在《著作权法》中制定关于为人工智能训练进行复制的明示规定。有引入日本式的非享受目的例外，或引入欧盟式的TDM例外和opt-out体系的方案。是否规定人工智能经营者有公开训练数据出处与构成的义务。是否保障著作权人确认自己的著作物是否被用于人工智能训练并拒绝使用的权利。截至2026年7月，其中尚无任何一项完成立法。

韩国《人工智能基本法》第2条第12项的“训练数据”定义与《著作权法》第35条之5的合理使用条款之间，存在着广阔的空白空间。一边广泛地定义了数据却未提及著作权，另一边仅规定了合理使用标准却未制定关于人工智能训练的具体规定。

如果首尔法院就广播公司诉门户网站经营者案做出判决，其事实上将代替立法。如果宽泛地认可合理使用，爬取将会继续。如果狭义地解释，所有训练数据都将需要许可。这是一种由法院的一份判决决定整个产业的数据获取成本的状况。

当《纽约时报》的辩护律师在诉状上写下“数十亿美元”时，这不仅仅是金额的大小。将人工智能训练数据视为免费原料，还是有偿许可的对象。对于那个问题，韩国的答案尚未得出。

第15章. 与《个人信息保护法》的连接点

2024年的某一天，韩国国内某大型平台企业的数据团队向公司内部的法务团队提出了一个质询：“我们打算将社交媒体上公开的数百万条帖子用于AI模型学习。这是使用者设置为公开的文字，还需要获得单独同意吗？”法务团队的回答只有两句话：“根据《个人信息保护法》，原则上需要获得同意。但是，《人工智能基本法》在第2条第12项中单独定义了学习用数据，却对个人信息处理只字未提。”这两句话之间，存在着实务上的空白。

1. 两部法律的关系：是一般法，还是并行？

《人工智能基本法》第5条第1款这样写道：“关于人工智能等，除其他法律有特别规定外，依照本法规定。”《个人信息保护法》第6条第1款的结构也几乎相同：“关于个人信息的处理及保护，除其他法律有特别规定外，依照本法规定。”两部法律都声称自己是一般法。

在这种结构下，处理包含在AI学习数据中的个人信息问题该何去何从呢？《人工智能基本法》将学习用数据定义为“用于人工智能开发和应用等的数据”（第2条第12项）。尽管法律本身认识到该数据中可能包含个人信息，但并未规定个人信息处理的合法性要件。

《个人信息保护法》规制个人信息处理的全过程。无论是AI学习还是统计分析，收集和利用能够识别特定在世个人的信息的行为，都是该法的适用对象。韩国法制研究院在2025年12月发布的报告《提高人工智能开发中个人信息处理合法性的方案》中分析称，《人工智能基本法》实际上将个人信息处理问题交给了《个人信息保护法》。

《人工智能基本法》第34条第3款创造了一个连接点：“人工智能经营者依照其他法令，按照总统令的规定履行了相当于第1款各项的措施时，视为履行了第1款规定的措施。”履行了《个人信息保护法》规定的安全措施和保障信息主体权

利义务的经营者，在个人信息处理和保护的范围内，将被认可为履行了相应的高影响人工智能责任。这是减轻重复规制的机制，但方向是单向的。遵守《个人信息保护法》可以满足《人工智能基本法》的义务，但遵守《人工智能基本法》并不能免除《个人信息保护法》的义务。

总结如下：对于处理AI学习数据中的个人信息，《个人信息保护法》是一般法，而《人工智能基本法》没有设立单独的特别规定。

2. 学习数据中的个人信息：收集与利用的同意问题

要训练AI模型，就需要数据。文本、图像、语音、视频。如果这些数据中包含个人的姓名、面部、声音、位置或行为模式，那就是个人信息。《个人信息保护法》第2条第1项将个人信息定义为“关于在世个人的信息，是通过姓名、居民注册号及影像等能够识别个人的信息”。即使仅凭该信息无法识别，但如果能与其他信息轻易结合并识别出个人，那也属于个人信息。

《个人信息保护法》第15条第1款列举了收集和利用的七种合法依据。包括信息主体的同意（第1项）、法律的特别规定（第2项）、合同履行（第4项）、正当利益（第6项）等。在以AI学习为目的的大规模数据采集中，想要获得数十万、数百万信息主体的个别同意，在现实中几乎是不可能的。抓取网络上公开的文本来训练大型语言模型的企业，能够征得所有原作者的同意吗？

在实务中，无需同意即可收集的依据主要讨论有两种：一是第15条第1款第6项的正当利益；二是第15条第3款的在与当初收集目的合理相关的范围内利用。

正当利益条款自2011年制定法律时就已存在：“为达成个人信息处理者的正当利益所必需，且明显优先于信息主体的权利的情况。”该条件十分苛刻。必须与正当利益有相当的关联性，且不得超出合理范围。能够主张AI学习的利益“明显”优先于信息主体隐私权的情况并不多。

第15条第3款关于合理相关范围内利用的规定，是将已合法收集的个人信息用于与当初目的不同之目的时的依据。使用者在利用服务过程中提供的数据，原本是出于服务运营目的收集的，后来转用于AI模型学习，便有可能适用此项。此时需要考虑是否会对信息主体造成不利，是否采取了确保安全性的措施等。

向第三方提供则更为严格。《个人信息保护法》第17条第1款原则上要求在向第三方提供时征得信息主体的同意。AI开发经营者从持有数据的企业接收学习用数据的行为，属于向第三方提供。作为无需同意即可提供的依据，有“在收集目的范围内提供”（第17条第1款第2项），但“提供数据用于AI学习”是否属于当初收集目的的范围，几乎总是存在争议。在与当初收集目的合理相关的范围内的提供，在第17条第4款中另有依据。在这种情况下，也必须考虑是否会对信息主体造成不利，是否采取了加密等确保安全性的措施。

3. 公开个人信息的陷阱

社交媒体个人资料、博客帖子、YouTube视频、企业主页的员工介绍。这些都是信息主体自行公开的信息。能够将这些信息用于AI学习吗？

《个人信息保护法》并没有针对“公开的个人信息”设立单独的条款。无论是否公开，只要是个人信息，就适用本法。信息主体在社交媒体上发帖，是向该平台使用者公开阅读的行为，而不是同意第三方企业将其用作AI模型学习数据的行为。

个人信息保护委员会在2024年7月发布了《为人工智能（AI）开发与服务处理公开个人信息的指南》。该指南认为，将公开的个人信息用于AI学习时，第15条第1款第6项的正当利益可以作为合法依据，但以采取安全措施和保障信息主体权利为条件。相应的组织架构也已搭建完毕。现行的个人信息保护委员会职务施行规则在个人信息政策局内设立了人工智能隐私团队。这是专门负责AI时代个人信息问题的部门。

欧盟在这个问题上采取了更为明确的立场。关于《欧盟一般数据保护条例》（GDPR）第6条第1款第(f)项的正当利益条款能否适用于AI学习，各国监管机构正在做出判断。意大利个人信息保护机构（Garante）在2023年3月要求OpenAI暂停了ChatGPT在意大利的服务。其判断是没有合法的处理依据。OpenAI随后提出了正当利益作为依据，但关于大型语言模型的网络爬虫行为能否以正当利益来正当化，相关争论并未结束。

在韩国法律中，结论也指向相同的方向。即使是公开的个人信息，也不能直接用于以AI学习为目的的大规模抓取。收集目的的合法依据、对信息主体的通知以及采取安全措施的义务依然如影随形。

4. 假名信息与匿名信息：AI学习的合法途径

《个人信息保护法》通过2020年数据三法的修订引入了假名信息制度。该制度为AI学习数据开辟了合法途径。

假名处理是指通过删除或替换个人信息的部分内容等方法，使其在没有附加信息的情况下无法识别特定个人的处理方式（第2条第1项之2）。假名信息仍然是个人信息的下位类别。因为它仍然保留了恢复到原始状态的可能性。然而，第28条之2第1款赋予了假名信息特例：“个人信息处理者为统计编制、科学研究、公益性记录保存等目的，可以未经信息主体同意处理假名信息。”

AI模型学习是否属于“科学研究”？《个人信息保护法》第2条第8项将科学研究定义为“技术的开发与实证、基础研究、应用研究及民间投资研究等应用科学方法的研究”。民间企业的AI模型开发也有很大余地被视为技术的开发与实证。在2020年修订当时的韩国国会讨论中，为了激活数字经济，也采纳了包含产业研究在内的广义定义。

使用假名信息虽可未经同意进行处理，但存在条件。不得以识别特定个人为目的的进行处理（第28条之5第1款）。如果在处理过程中生成了能够识别个人的信息，必须立即停止并予以销毁（同条第2款）。同时也被赋予了对假名信息采取安

全措施的义务（第28条之4）。如果要在不同的个人信息处理者之间结合假名信息，必须经过个人信息保护委员会或有关中央行政机关指定的专门机构（第28条之3）。

假名信息制度的局限性也很明显。大型语言模型学习的是整个文本。仅对名字进行假名处理并不能解决问题。如果上下文中保留着能够识别个人的描述，那么该文本依然是个人信息。

匿名信息则有所不同。《个人信息保护法》第58条之2写道：“在合理考虑时间、成本、技术等因素时，即使使用其他信息也无法再识别出个人的信息”，不适用本法。完全匿名化的数据不属于个人信息。由于不受《个人信息保护法》适用，因此既不需要同意，也不需要安全措施，且没有处理限制。可以不受限制地用于AI学习。问题在于，“完全匿名化”的标准很高。在大型数据集中重新识别个人的技术仍在不断发展。

对经营者而言，现实的途径可以概括如下：若想在未经同意的情况下使用包含个人信息的学习数据，要么在假名处理后出于科学研究目的进行处理，要么进行完全匿名化，从而脱离《个人信息保护法》的适用。无论哪种方式，都需要技术措施和法律验证的双管齐下。

5. 两种要求说明权

《人工智能基本法》和《个人信息保护法》分别规定了“获得说明的权利”。两者条文的结构和对象各不相同。

《人工智能基本法》第3条第2款规定：“受影响者对于推导出人工智能最终结果所运用的主要标准及原理等，在技术上和合理可能的范围内，应当能够获得明确且有意义的说明。”该条款被列入基本原则规定中。并未规定具体程序或违反时的制裁措施。“受影响者”是指因人工智能产品或服务而在生命、身体安全或基本权利方面受到重大影响的人（第2条第9项）。这并非指所有的AI使用者，仅限于受到重大影响的人。

《个人信息保护法》第37条之2则有所不同。该条款通过2023年的修订而新设，并于2024年3月15日施行：“信息主体在由完全自动化的系统（包括应用了人工智能技术的系统）处理个人信息所做出的决定（排除《行政基本法》第20条规定的行政厅的自动处分），对其权利或义务产生重大影响的情况下，有权向相关个人信息处理者拒绝该决定。”该条款同时赋予了要求说明权（第2款）和拒绝权（第1款）。

通过比较两者差异，轮廓便清晰可见。

《人工智能基本法》的要求说明权赋予“受影响者”。对象是普遍的人工智能系统。无论是否具有高影响性，只要是受到重大影响的人即可。然而，它仅停留在基本原则条款层面，缺乏履行程序。《个人信息保护法》第37条之2则赋予“信息主体”。对象是“由完全自动化的系统处理个人信息所做出的决定”。信息主体可以拒绝该决定，也可以要求说明，而个人信息处理者除非有正当理由，否则必须采取通过人为介入进行重新处理等必要措施（第3款）。此外，还负有公开自动化决定的标准、程序以及个人信息处理方式的义务（第4款）。

假设某人被AI贷款审批系统拒绝了贷款。此人既是《人工智能基本法》上的受影响者，同时也是《个人信息保护法》上的信息主体。两部法律的要求说明权同时适用。在实务中，经营者应依据哪部法律提供何种程度的说明，以及两部法律的要求发生冲突时哪一方优先，均不明确。由于《个人信息保护法》第37条之2具备了履行程序和制裁措施，因此这一侧的实质约束力更强。

6. 用户画像与自动化决定

用户画像（Profiling）是为了分析或预测个人的特定层面而自动处理个人信息的行为。信用分计算、客户行为分析、定向广告投放等均属此类。AI恰好就是在做这项工作。

《个人信息保护法》第37条之2中“由完全自动化的系统处理个人信息所做出的决定”包含了基于AI的用户画像结果。该条款附有但书。当自动化决定获得了

信息主体同意（第15条第1款第1项）、为遵守法定义务所不可避免（第2项）或为履行合同所必需（第4项）时，拒绝权不适用。

这种结构类似于欧盟GDPR第22条。GDPR第22条第1款规定：“信息主体有权不受包含用户画像在内、对其产生法律效力或与之类似重大影响的、仅基于自动化处理做出的决定之约束。”这是原则性的禁止。例外情况仅限于为履行合同所必需、联盟法或成员国法律允许，或信息主体明确同意的情况（第22条第2款）。

韩国法律与GDPR的差异在于出发点。GDPR原则上禁止自动化决定，并开辟了例外。韩国《个人信息保护法》并不禁止自动化决定，但赋予信息主体拒绝权和要求说明权。两者的处理方式不同。欧盟是“原则上禁止+例外允许”，而韩国是“允许+拒绝权+说明权”。

《人工智能基本法》并未使用用户画像（Profiling）一词。但是，第2条第4项第(7)目将“招聘、贷款审批等对个人权利和义务关系产生重大影响的判断或评估”列为高影响人工智能的适用领域。招聘筛选、贷款审批、保险风险评估。在这些领域，如果AI通过分析个人数据做出判断，那么《人工智能基本法》规定的高影响人工智能义务（风险管理、说明、人为介入、文件化）和《个人信息保护法》对自动化决定的规制将同时适用。

假设金融公司使用AI进行贷款审批。根据《人工智能基本法》，必须制定风险管理方案，说明主要标准，并建立由人员进行管理和监督的体系。根据《个人信息保护法》，必须公开自动化决定的标准和程序，且当被拒绝的申请人要求说明时，必须进行人为介入的重新处理。甚至还会叠加《信用信息法》中有关自动化评估的条款。一项服务将重叠适用三部法律。

7. 两种影响评估

《人工智能基本法》第35条规定了高影响人工智能影响评估：“人工智能经营者提供使用高影响人工智能的产品或服务时，应当努力在事前评估对人的基本权利产生的影响。此时，影响评估应当考虑使用高影响人工智能的产品或服务的性

质，以便反映出人工智能弱势群体的特性。”这是一项自主努力的义务。国家机关如果想使用高影响人工智能产品，应当优先考虑已实施影响评估的产品（第2款）。

《个人信息保护法》第33条规定了个人信息影响评估。公共机关的首长在运营符合总统令规定标准的个人信息文件，担忧会侵害信息主体个人信息时，应当进行影响评估，并将结果提交给保护委员会。这对公共机关而言是强制性义务。民间的个人信息处理者则承担“应当积极努力”的自主努力义务（第11款）。

这两种影响评估名称相似，但评估对象不同。AI影响评估关注的是AI系统对人的基本权利所产生的影响。评估项目包括偏见、歧视、安全性、透明度等AI固有的风险。个人信息影响评估则关注个人信息文件的运营对侵害信息主体个人信息所产生的影响。评估项目包括处理的个人信息数量、是否向第三方提供、侵害信息主体权利的可能性等。

如果高影响人工智能处理个人信息，公共机关可能需要实施这两种影响评估。引入AI招聘系统的公共机关，必须分别执行根据《人工智能基本法》进行的AI影响评估，以及根据《个人信息保护法》进行的个人信息影响评估。两者既有评估项目重合的部分（歧视与偏见、侵害信息主体权利），也有不重合的部分（AI的技术安全性、个人信息的保存期限）。

《人工智能基本法》第34条第3款的缓解重复规制条款并不涉及此处。因为该履行视为条款仅适用于第34条第1款的经营者责任，而第35条的影响评估是独立的条文。即使完成了个人信息影响评估，由于该评估并未涵盖AI固有的风险（偏见、不透明性、易受对抗性攻击等），因此AI影响评估依然是需要单独努力的课题。

8. 与欧盟GDPR第22条的距离

欧盟GDPR第22条是对韩国法律产生影响的先行规范。2023年通过修订引入《个人信息保护法》第37条之2时，在立法讨论中便将GDPR第22条作为参考模型

进行了探讨。然而，这两项条款的结构和效果之间存在差距。

GDPR第22条以“仅基于自动化处理做出的决定”（decisions based solely on automated processing）为对象。有人员进行实质性介入的决定不在此列。韩国《个人信息保护法》第37条之2则以“由完全自动化的系统处理个人信息所做出的决定”为对象。虽然“完全”这一表述与GDPR的“solely”相对应，但究竟多大程度的人类介入才算脱离了“完全的自动化”，双方都留有解释的余地。

欧盟《人工智能法案》（EU AI Act）在GDPR第22条之上叠加了新的层级。附件III中的高风险人工智能系统涵盖了在就业、教育、必要服务、信贷获取等领域使用的AI，并且对此类AI强制要求进行符合性评估、技术文件化、日志保存以及人为监督。在GDPR第22条的拒绝权与《人工智能法案》的符合性评估发生重叠的领域，欧盟法律将同时适用这两部法律。韩国也是同样的结构。《个人信息保护法》第37条之2与《人工智能基本法》第34条将同时适用。

显现差异的痛点在于制裁的力度。违反GDPR的罚款最高可达2000万欧元或全球年营业额的4%。韩国《个人信息保护法》虽有罚款规定（第64条之2），但《人工智能基本法》的最高行政罚款仅为3000万韩元。这一差距将直接影响经营者合规的动力。

9. 实务清单：经营者应确认的五件事项

梳理开发或使用AI的经营者在《个人信息保护法》与《人工智能基本法》交叉领域中需要确认的事项。

学习数据中是否包含个人信息。如果文本、图像、语音、视频数据中含有能够识别个人的信息，便适用《个人信息保护法》。在未进行假名处理或匿名化的情况下，未经同意的收集具有很高的违法风险。

是否进行了假名处理或匿名化。假名信息可以出于科学研究目的未经同意进行处理，但必须遵守禁止重新识别的义务和采取安全措施的义务。匿名信息虽然不

受《个人信息保护法》的适用，但经营者有责任证明匿名化的充分性。

AI系统是否做出自动化决定。如果该系统做出的自动化决定对个人的权利或义务产生重大影响，则同时适用《个人信息保护法》第37条之2的拒绝权和要求说明权，以及《人工智能基本法》第3条第2款的要求说明权。必须建立人为介入的体系。

高影响人工智能是否处理个人信息。如果在高影响人工智能领域（招聘、贷款审批、医疗、公共服务等）处理个人信息，那么《人工智能基本法》规定的高影响人工智能义务将与《个人信息保护法》规定的安全措施及影响评估义务相重叠。根据《人工智能基本法》第34条第3款，遵守《个人信息保护法》可以部分满足《人工智能基本法》的义务，但仍需单独履行AI固有的义务（风险管理、透明度）。

是否抓取公开的个人信息用于学习。即便是信息主体公开的信息，以AI学习为目的进行大规模收集也需要单独的合法依据。即便援引正当利益条款（第15条第1款第6项），也很难证明其明显优先于信息主体的权利。

10. 法律的空白地带

《人工智能基本法》对个人信息的处理保持沉默，而《个人信息保护法》尚未制定符合AI学习这一新现实的特别规定。假名信息制度虽然提供了一条途径，但在大型语言模型学习整个网络内容的现实面前，“假名处理后用于科学研究目的”这条路径显得过于狭窄。

个人信息保护委员会发布指南并运营人工智能隐私团队，正是试图填补这一空白的举动。2026年3月10日公布的《个人信息保护法》修正案（法律第21445号，预计分别于2026年9月11日和2027年7月1日施行）似乎也未包含正面探讨AI学习的特别规定。符合AI时代的法律层面的特别规定，成为了下一次修订留下的课题。

《人工智能基本法》第5条第2款写道：“在制定或修改关于人工智能等的其他法律时，应当使其符合本法的目的。”《个人信息保护法》中有关AI的后续修订也将受到该条款的约束。两部法律将如何衔接，最终将由实务和执行来决定，而不是法律文本。目前，正是经营者的质询在两部法律之间不断堆积的时期。

第16章 行业邻接规制的重复

2026年3月，一家运营基于人工智能皮肤分析应用程序的初创公司的合规负责人，在法务团队会议上展示了一份Excel文件。纵轴列出了其服务的十二项功能，横轴排列着《人工智能基本法》、《医疗器械法》、《数字医疗产品法》、《个人信息保护法》、《标识广告法》、《电子商务法》六部法律。每个单元格里密密麻麻地写满了义务条款的编号。仅风险管理义务就重叠涉及三部法律，而使用者告知义务则被四部法律分别要求。负责人问道：“在这些法律中，我们应该首先遵守哪一部？”

《人工智能基本法》是水平的普通法。它不分产业领域，对所有使用人工智能的经营者赋予义务。然而，韩国在金融、医疗、通信、个人信息、消费者保护等各个领域，已经存在积累了数十年的垂直规制。如果水平的新法律重叠在现有的垂直规制之上，经营者将面临对同一行为要在两个以上的地方以不同格式制作相似文件的处境。韩国国会立法调查处（NARS）在《人工智能基本法》施行前就指出，有必要就金融、医疗、机器人等各领域的人工智能立法进行讨论，韩国法制研究院（KLRI）也警告称，在各领域人工智能规制成熟之前就率先制定基本法，导致了层级颠倒。

1. 第5条原则：特别法优先

《人工智能基本法》第5条由两个款组成。

第1款宣布：“关于人工智能等，除其他法律有特别规定外，依照本法的规定。”这是典型的“普通法-特别法”优先规定。如果《医疗器械法》对人工智能医疗器械设有单独条款，那么在该领域将优先适用《医疗器械法》，《人工智能基本法》则退居其次。

第2款规定：“在制定或修订关于人工智能等的其他法律时，应当符合本法的目的。”这意味着，即使以后制定了各领域的法律，也不能偏离《人工智能基本法

》的理念和方向。

原则很明确。如果现有的各领域法律中已经有规制人工智能的条款，则该法律优先。如果没有，《人工智能基本法》则补充适用。然而，原则与实务之间存在差距。因为对于“有特别规定的情况”具体指什么，以及现有法律的一般安全义务是否属于可以替代《人工智能基本法》高影响人工智能义务的“特别规定”，解释可能会出现分歧。

2. 施行令附表1：削减重复负担的利刃

《人工智能基本法》针对重叠问题准备了一个具体的解决方案。即第34条第3款：“人工智能经营者按照总统令的规定，根据其他法令履行了相当于第1款各项的措施时，视为履行了第1款规定的措施。”这意味着如果遵守了其他法律，就会被认可为也遵守了本法。

施行令第27条第5款即为该总统令：“根据法第34条第3款，视为人工智能经营者履行了同条第1款规定措施的情形如附表1所示。”附表1按领域逐条列出了履行哪部法律的哪个条款，就被视为履行了《人工智能基本法》第34条第1款第1项至第5项的五项义务（风险管理、说明、使用者保护、人为管理监督、文档化）中的哪一项。虽然第34条第1款还有要求履行委员会审议和表决事项的第6项，但附表1将承认履行的对象限定在第1项至第5项。

八个范畴被列入附表1。分别是数字医疗器械、自动驾驶汽车、《汽车管理法》规定的汽车网络安全、自主航行船舶、核物质与核设施、贷款审查、《智能化信息基本法》的安全保护措施、《个人信息保护法》的保护措施。

数字医疗器械在项目数量上占据压倒性优势。如果根据《数字医疗产品法》建立质量管理体系并获得质量管理标准合格判定，即视为履行了《人工智能基本法》的风险管理（第1项）义务。获得制造许可或进口许可并在软件上标注信息，则说明（第2项）义务也得到满足。如果履行了遵守事项，即视为履行了使用者保护（第3项）；配备质量负责人，即视为履行了人为管理监督（第4项）；建立

质量管理体系并获得合格判定，即视为履行了文档化（第5项）。这五项义务全部可以在《数字医疗产品法》体系内得到解决。

贷款审查领域也包含在内。如果遵守了《信用信息法》第35条之2规定的说明义务，并具备了第36条之2第1款规定的自动化评估说明程序，则视为履行了《人工智能基本法》的说明义务。如果履行了《金融消费者保护法》第10条规定的职责，则使用者保护义务也得到满足。

对于自动驾驶汽车，如果获得《自动驾驶汽车法》规定的性能认证和变更认证，则视为履行了风险管理（第1项）、使用者保护（第3项）、文档化（第5项）义务。人为管理监督（第4项）义务必须在履行了同法第43条规定的所有职责后才被认可，而对说明（第2项）义务则没有认定项目。对于自主航行船舶，如果获得《自主航行船舶法》规定的安全性评估和航行批准，则视为履行了全部五项义务。

《汽车管理法》规定的汽车也包含在附表1中。获得汽车网络安全管理体系认证和变更认证，并履行了同法第34条之5规定的所有职责，则视为履行了风险管理（第1项）义务。对于核物质与核设施，如果《关于核电站违规防范及核电事业者等的管理监督法律》规定的遵守事项或伦理行为准则中包含相关内容，则视为履行了说明（第2项）和人为管理监督（第4项）义务。

《个人信息保护法》被作为单独项目（附表1第7项）处理。如果履行了《个人信息保护法》第29条（安全措施义务）、第30条（个人信息处理方针）、第31条（个人信息保护负责人）、第35条（个人信息的阅览）、第36条（个人信息的更正删除）、第37条（个人信息的停止处理）等12个条款的措施和义务，“对于个人信息的处理及保护”，视为履行了《人工智能基本法》第34条第1款各项的措施。由于附带了“对于个人信息的处理及保护”这一限定语，因此与个人信息无关的风险管理义务并未被免除。

如果履行了《智能化信息基本法》第60条第1款第1项至第3项及第5项的安全性保护措施，则视为履行了风险管理（第1项）和说明（第2项）义务；如果履行了同条款第4项的保护措施，则视为履行了人为管理监督（第4项）义务。

3. 附表1留下的空白

有一些领域未在附表1中列出。

能源供应、饮用水生产、医疗保健体系建设、核能安全（附表1中仅包含部分项目）、生物识别信息分析、公共机构决策、学生评估。在《人工智能基本法》第2条第4项列举的高影响人工智能11个领域中，附表1制定了具体履行认定标准的仅有医疗器械、贷款审查、交通工具（自动驾驶汽车、汽车网络安全、自主航行船舶）、核能（部分）等。在其余领域开展业务的经营者必须同时遵守现有法令和《人工智能基本法》。

招聘领域就是一个典型例子。《招聘程序公平化法》并未对人工智能面试或人工智能简历筛选设定单独条款。虽然韩国雇佣劳动部在2023年12月发布了《人工智能招聘系统指南》，但这并不具有法律约束力。在招聘中使用人工智能的经营者必须全面履行《人工智能基本法》规定的高影响人工智能义务，同时还要遵守《招聘程序公平化法》的现有义务。由于附表1中遗漏了招聘，因此即使遵守了《招聘程序公平化法》，也不能免除《人工智能基本法》的义务。

学生评估也是如此。《教育基本法》中没有关于基于人工智能的评估规定。利用人工智能进行大学修学能力考试评分或学生成绩计算，将直接适用《人工智能基本法》的高影响人工智能义务，但在韩国教育部制定单独的指南或公告之前，履行方法依然模糊不清。

4. 《标识广告法》：一则广告适用两部法律

我们来看将一张人工智能制作的广告图片上传到互联网的情况。

《人工智能基本法》第31条第2款要求对生成式人工智能生成的成果物标识“由生成式人工智能生成的事实”。第3款要求对难以与真实区分的虚拟图像或视频，以“使用者能清晰识别的方式”进行告知或标识。

《关于标识广告公平化的法律》（《标识广告法》）第3条第1款禁止“存在欺骗消费者或使消费者误解之虞”的标识和广告。虚假、夸大或欺骗性的标识广告均属此列。如果使用人工智能生成的模特图像制作化妆品广告，却表现得像实际使用效果一样，就可能构成《标识广告法》上的欺骗性广告。

两部法律的义务发生了重叠。《人工智能基本法》要求标识“由人工智能生成”这一事实本身，而《标识广告法》则要求“不得欺骗消费者”。方向虽然一致，但标识方法却不同。履行了《人工智能基本法》的标识义务，并不意味着自动满足了《标识广告法》禁止不当广告的义务。反之亦然。即使写明“本图像由人工智能生成”，如果该广告的内容本身属于夸大，依然违反《标识广告法》。

应用第5条的逻辑，《标识广告法》是关于广告真实性的“特别规定”，因此在判断广告内容是否具有欺骗性时，《标识广告法》优先。《人工智能基本法》的人工智能生成物标识义务与《标识广告法》在不同的层面上运作。结果，经营者必须同时遵守这两部法律。附表1中并未包含《标识广告法》。

在韩国公平交易委员会尚未出台关于人工智能生成广告的审查指南的情况下，关于标识方法、位置和大小的具体标准，在这两部法律中均未得到充分规定。作为经营者，只能将《人工智能基本法》的确保透明度指南与《标识广告法》的现有审查指南并列放置，自行整理重叠的部分。

5. 《数字医疗产品法》：紧密相连的领域

与《人工智能基本法》像齿轮般精密咬合的法律是《数字医疗产品法》。

《数字医疗产品法》于2024年1月23日公布，并于2025年1月24日施行，部分条款从2026年1月24日起施行。《人工智能基本法》第2条第4项d目直接指名了

《数字医疗产品法》。该条款明确指出，“开发及使用《医疗器械法》规定的医疗器械及《数字医疗产品法》规定的数字医疗器械”是高影响人工智能的领域之一。该目中关于数字医疗器械的部分于2026年1月24日施行，紧随《人工智能基本法》施行日（2026年1月22日）两天后生效。

我们来看制造基于人工智能的医疗器械的经营者的义务。由韩国食品药品安全处（食药处）管辖的《数字医疗产品法》对制造许可、质量管理体系、临床试验、软件信息标识、选任质量负责人，直至上市后安全管理的全过程进行规制。《人工智能基本法》则要求风险管理、说明、使用者保护、人为管理监督、文档化。

得益于附表1，忠实遵守《数字医疗产品法》体系的经营者的义务，无需额外履行《人工智能基本法》规定的高影响人工智能义务。因为针对附表1中的五项义务，均已指定了《数字医疗产品法》的对应条款。

但是有一点需要注意。附表1免除的仅是《人工智能基本法》第34条（高影响人工智能经营者职责）规定的义务。第31条（确保透明度义务）的事先告知和生成物标识、第35条（影响评估）的努力义务并未包含在附表1的适用范围内。即使是人工智能医疗器械经营者，也必须另外向使用者事先告知“本服务基于人工智能运行”。

欧盟《人工智能法案》（EU AI Act）对医疗器械采取了不同的处理方式。欧盟的《医疗器械法规》（MDR）和《体外诊断医疗器械法规》（IVDR）属于《人工智能法案》附件一（Annex I）中列出的产品安全规制。根据这些法规接受符合性评估的医疗器械，将依据《人工智能法案》第6条第1款自动分类为高风险人工智能，且符合性评估程序被合二为一。韩国采取的不是整合，而是相互认定方式。在承认遵守现有法令即视为履行的同时，两部法律仍然保持独立的体系。

6. 金融：在指南与法律之间

金融领域的人工智能监管格局十分复杂。

韩国金融委员会在2021年7月发布了《金融领域人工智能指南》，并在2024年12月出台了金融人工智能七大原则。该指南是不具有法律约束力的行政指导。它建议对人工智能信用评估模型的公平性、可解释性和数据质量进行管理。

《信用信息法》第36条之2以法律形式规定了对自动化评估的说明义务和异议提出权。作为个人的信用信息主体有权听取关于自动化评估结果、主要标准、基础信息概要的说明，如果基础信息不准确，可要求更正或删除，或要求重新计算结果。该条款在《人工智能基本法》制定之前就已经存在。

《人工智能基本法》第34条第1款第2项要求高影响人工智能经营者制定并实施关于“得出最终结果所采用的主要标准”等方面的说明方案。在贷款审查中使用人工智能的金融公司，将同时承担《信用信息法》的说明义务和《人工智能基本法》的说明义务。

附表1解决了这个问题。如果遵守《信用信息法》第35条之2规定的说明义务，并具备了第36条之2第1款规定的程序，则视为履行了《人工智能基本法》的说明义务。如果履行了《金融消费者保护法》第10条规定的职责，则使用者保护义务也得到满足。

问题在于贷款审查以外的金融人工智能。保险承保审查、异常交易检测（FDS）、资产管理智能投顾、内部信用等级测算。这些可能被纳入《人工智能基本法》第2条第4项g目中“对招聘、贷款审查等个人权利义务产生重大影响的判断或评估”，但附表1仅明确指出了“贷款审查”。在保险或投资中使用人工智能的金融公司，即使遵守了现有的金融规制，是否就能自动免除《人工智能基本法》的义务尚不明确。

仅遵循韩国金融委员会的人工智能指南，并不能被认可为履行了《人工智能基本法》义务。因为附表1仅将法律和施行令级别的义务作为相互认定的对象，而指南属于没有法律约束力的行政指导。金融公司必须同时履行韩国金融委员会指南和《人工智能基本法》义务的双重负担依然存在。

7. 《个人信息保护法》与《信用信息法》：重叠层最厚的地方

人工智能系统几乎无一例外地会处理个人信息。训练数据中包含个人信息，推理结果也包含了对个人的判断。实际上，《人工智能基本法》和《个人信息保护法》将同时适用于所有人工智能经营者。

附表1第7项试图综合处理这一问题。列举了《个人信息保护法》的12个条款，只要履行了这些条款，“对于个人信息的处理及保护”，即视为履行了《人工智能基本法》第34条第1款各项的措施。其中包括安全措施义务（第29条）、公布处理方针（第30条）、指定保护负责人（第31条）、阅览权（第35条）、更正删除权（第36条）、停止处理权（第37条）等。

范围限定语“对于个人信息的处理及保护”是关键。在人工智能系统的风险中，只有与个人信息相关的部分才会被免除。对于因算法偏见导致的歧视、因无法解释的判断导致的基本权受到侵害、安全事故风险等不与个人信息直接相关的风险，《人工智能基本法》规定的义务依然有效。

韩国法制研究院于2025年12月发布的金度承教授的报告《提高人工智能开发中个人信息处理合法性的方案》，正面分析了这一冲突点。其核心指出，在训练数据收集阶段，《个人信息保护法》的同意原则与人工智能开发的现实之间存在脱节。如果要用数万份医疗影像来训练人工智能，在《个人信息保护法》的同意体系下，有时在现实中几乎是不可能的。《人工智能基本法》虽然对训练用数据设有单独规定，但与《个人信息保护法》的关系尚未完全理清。

《信用信息法》在金融领域扮演着《个人信息保护法》特别法的角色。其中规定了自动化评估说明义务（第36条之2），并对假名处理和匿名处理设有单独规定。进行人工智能贷款审查的金融公司处于《信用信息法》、《个人信息保护法》和《人工智能基本法》三部法律的交叉点。虽然附表1允许以《信用信息法》的说明义务替代《人工智能基本法》的说明义务，但《个人信息保护法》中关于同意、阅览、更正和删除的义务在不同层面上依然适用。

8. 防止重复负担的机制能发挥到何种程度

我们再次审视附表1的结构。

附表1所免除的仅限于《人工智能基本法》第34条第1款规定的五项义务（风险管理、说明、使用者保护、人为管理监督、文档化）。第31条（确保透明度义务）的事先告知和生成物标识、第32条（确保安全性义务）的高性能人工智能风险管理和结果提交、第33条（确认高影响人工智能）的事先研讨、第35条（影响评估）的努力义务，均不在附表1的适用范围内。换句话说，即使是完全遵守了《数字医疗产品法》的人工智能医疗器械经营者，也必须独立履行《人工智能基本法》的透明度告知义务和影响评估努力义务。

施行令第27条第3款也是减少重复负担的机制。如果前置的人工智能经营者已经履行了第34条第1款第1项至第3项（风险管理、说明、使用者保护）的措施，提供该人工智能系统的使用经营者只要“没有发生显著改变其本来目的或用途等重大功能变更”，即视为履行了相同的措施。如果前置的人工智能经营者已经进行了风险管理，使用经营者就不必再做一次。然而，如果使用经营者将人工智能系统应用到其他语境下并进行了有意义的更改，则不能适用此项豁免。

9. NARS的警告：需要进行各领域的立法

韩国国会立法调查处立法调查官郑俊华在2025年3月发布的《人工智能基本法实施前补充立法课题》中，提出了实施前需要完善的课题。其中之一就是有必要讨论金融、医疗、机器人等各个领域的人工智能立法。

韩国法制研究院研究委员郑元俊在2025年9月的报告《人工智能基本法的合理规制及发展方案研究》中指出了更为根本的问题。虽然《人工智能基本法》冠以“基本法”之名，但基本法原本的功能是为下位个别法设定方向。通常的顺序是先制定个别法，之后基本法再将其囊括其中；而韩国是在各领域人工智能法律几乎缺失的状态下，率先出台了基本法。在个别法处于空白的状态下，基本法直接向经营者赋予义务，这导致了与现有的各领域规制产生尴尬重叠的情况。

郑元俊研究委员建议：应按照高影响人工智能的细分领域，由相关领域的主管部门整理有关人工智能的规定，并明确界定与《人工智能基本法》的关系。目前的附表1只是一个开始，应当扩大到招聘、公共服务、教育、能源、生物识别等尚未包含在附表中的领域。

10. 欧盟的方法：将人工智能嵌入现有的规制中

欧盟的《人工智能法案》从一开始就在设计上反映了与现有规制之间的关系。

对于医疗器械（MDR/IVDR）、机械设备（Machinery Regulation）、汽车一般安全（General Safety Regulation）、玩具安全、海事设备等符合欧盟产品安全协调立法（列入附件一）的产品，如果人工智能作为安全组成部分搭载在其中，将在该产品原有的符合性评估程序内，一并审查《人工智能法案》规定的高风险人工智能要求。这不是进行两次评估，而是整合为一个程序。

韩国并未采取这种整合模型。《数字医疗产品法》的许可程序和《人工智能基本法》的义务履行作为各自独立的程序存在，但只要履行了前者，就“认定”为履行了后者。虽然实际结果可能相似，但法律结构却不同。欧盟以一份证书即可完成，而韩国则需要分别确认这两部法律的履行与否。

欧盟的数字一揽子法案（Digital Omnibus）修正案进一步推动了这种整合。在将高风险人工智能义务的适用期限推迟至2027年12月（独立型）和2028年8月（产品内置型）的同时，也为确保与现有产品安全规制的一致性争取了时间。同时还提出了将中小企业合规成本降低35%的目标。

中国则又有不同。由于没有全面的人工智能法律，而是优先制定了各领域的规定，因此发生重叠问题的余地本身较小。因为《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》以及《生成式人工智能服务管理暂行办法》各自狭义地指定了适用对象。不过这种做法是有代价的。当一个人工智能服务同时适用多项规定时，缺乏确定哪项规定优先的一般原则。中国正在准备综合性人工智能法的原因之一，正是为了整合这种碎片化的监管体系。

11. 经营者在实务中应遵循的判断顺序

对于询问应当优先遵守哪部法律的经营者，我们建议采用以下顺序。

首先，确认本公司的人工智能服务是否已经属于某个特定领域的规制之中。比如是获得了医疗器械许可的产品，还是获得了金融业许可而运营的服务，亦或是获得了自动驾驶性能认证的车辆。如果存在各领域的规制，该规制就是首要的遵守对象。根据《人工智能基本法》第5条第1款，“其他法律有特别规定的情况”，该法律优先适用。

其次，确认施行令附表1。如果本公司已履行的各领域法律条款列在了附表1中，则无需额外履行《人工智能基本法》第34条相应的义务。如果已经具备了《数字医疗产品法》规定的质量管理体系，就满足了《人工智能基本法》的风险管理义务。

接着，确认未包含在附表1中的《人工智能基本法》义务。事先告知透明度（第31条）、确认高影响人工智能（第33条）、影响评估努力义务（第35条），均必须独立于附表1进行履行。

最后，如果本公司人工智能服务属于高影响人工智能11个领域且不在附表1包含的领域内（如招聘、公共服务、学生评估、能源等），则必须全面履行《人工智能基本法》第34条的义务。即便遵守了现有的特定领域法律，也无法获得豁免。

即使遵循了上述顺序，仍会遇到模棱两可的情况。此时可以向人工智能基本法支持中心（支持台）咨询。实施第10周共受理了552件咨询，从在线咨询数据来看，排名前列的项目是“透明度”（56.4%）和“确认高影响人工智能”（17%），这一事实表明了经营者所面临的困惑集中在何处。

12. 监管版图的完成条件

《人工智能基本法》施行令附表1是防止重复负担的第一版。数字医疗器械、自动驾驶汽车、汽车网络安全、自主航行船舶、核物质与核设施、贷款审查、《智

能化信息基本法》、《个人信息保护法》。共有八个范畴名列其中。其余部分还是空白。

经历了2026年5月21日的立法预告并于7月20日公布的修订施行令中，同样未包含对附表1的扩充。仍需等待韩国金融委员会、雇佣劳动部、教育部、环境部、国土交通部等各领域的主管部门整理其所在领域的人工智能规制，并正式界定与《人工智能基本法》的关系。在完成该工作之前，经营者必须在两部法律的交叉点上自行作出判断。

空白部分被填补的速度越快，经营者的合规成本就越低。空白存留的时间越长，对同一行为必须向多处提交不同文件的状况就会持续越久。清除会议室Excel文件中的重复标记，并非法律人士的职责，而是立法的职责。

第六部 经营者可以获得的支持与机会

第17章. 学习用数据、基础设施、集聚园区，获得政府支持的方法

一家试图创建胸部X光判读模型的人工智能初创公司碰壁了。虽然需要10万张医疗影像数据，但得到的答复是，仅在各个医院接受IRB审查并签订数据使用合同就需要8个月。对于一家只有12名员工的公司来说，8个月是资金耗尽的时间。《人工智能基本法》将学习用数据综合提供系统写入法律的原因就在于此。

1. 学习用数据综合提供系统的结构

法律第15条第4款规定，科学技术信息通信部长官应当建立和管理能够综合提供和管理学习用数据的系统，即综合提供系统，并提供给民间自由使用。施行令第13条勾勒了该系统的具体轮廓。

施行令第13条第1款规定的系统功能包括学习用数据的综合搜索、系统分类、出处提供，与国家机关和民间运营的其他数据平台的联动，以及提供价值评估和质量管理相关信息。经营者在该系统中搜索并使用所需的数据集。

科学技术信息通信部长官建立并管理系统。法律第39条第2款规定了可以将综合提供系统的建立、运营和管理业务委托给专业机构的依据。韩国智能信息社会振兴院（NIA）一直运营的AI Hub（aihub.or.kr）已经发挥了类似的作用，而第15条第4款以综合提供系统的名义确立了直接的法律依据。

使用综合提供系统并不总是免费的。根据法律第15条第5款和施行令第14条，科学技术信息通信部长官可以向使用者收取费用，并可以根据学习用数据的种类和使用目的实施差别化的使用费。公共机构使用或非营利研究机构、教育机构出于学术研究、教育目的使用时，可以减免使用费。征收标准和减免对象的细节由科学技术信息通信部长官通过公告规定。

2. 公共数据开放与提供人工智能学习目的

《公共数据法》（关于促进公共数据提供和使用的法律）自2013年起施行。然而，该法律并不是以人工智能学习为前提制定的。未整理成标准化格式的公共数据很难直接用于人工智能学习。

《人工智能基本法》第16条第2款第2项之2允许国家和地方自治团体提供必要的支持，以便将公共机构持有和管理的数据生成并提供为学习用数据，并确保适当的质量水平。施行令第13条第2款规定，科学技术信息通信部长官可以向中央行政机关、地方自治团体及公共机构的负责人请求协助，如综合提供系统与各机构的系统、数据的联动，提交资料等，被请求的机构负责人除非有特殊事由，否则应当遵从。将公共数据作为学习用数据提供的事项需与行政安全部长官协商（施行令第13条第3款）。这超越了现有《公共数据法》的一般开放义务，在政府方面开辟了将适合人工智能学习形式的加工数据输入系统的道路。请求的主体是科学技术信息通信部长官，并不是赋予经营者向政府请求加工数据的权利。

在实务中，这产生了巨大的差异。以气象厅的天气数据为例。从公共数据门户下载的CSV文件存在许多缺失值，并且地区代码体系每年都在变化。若要用于人工智能学习，需要进行缺失值校正、代码整合和时间序列排序等工作，2名工程师有时需要花费3周时间进行这种预处理。如果通过综合提供系统获得为学习目的加工的数据集，就能省去这3周时间。

但也存在局限性。并非所有公共数据都会立即提供用于人工智能学习。涉及国家安全、个人信息和第三方著作权的数据被排除在外。经营者必须首先确认所需的数据集是否已在综合提供系统中注册。如果没有所需的数据，可以向运营机构提出需求，但这并非法律保障的请求程序，而只是实务上的要求。

3. 数据质量管理支持

法律第15条第1款规定，科学技术信息通信部长官应当与相关中央行政机关的负责人协商，推进确保学习用数据质量水平等所需的政策。施行令第13条第4款赋予了科学技术信息通信部长官努力保持通过综合提供系统提供的学习用数据的

最新性、准确性和相互关联性的义务。

这对经营者而言不是义务，而是福利。因为确保质量的责任落在了政府一方。不过，《人工智能基本法》中没有数据质量认证制度。数据集质量认证是《数据产业法》体系下单独的制度，因此需要认证的经营者必须单独确认该法制的程序。

为什么质量标准很重要呢？因为人工智能模型的性能直接取决于数据质量。用掺杂标签错误的数据学习的模型会原封不动地学到这些错误。在很多情况下，提高数据质量比增加数据量更具成本效益。

根据法律第15条第2款，政府可以选定关于学习用数据生产、收集、管理、流通和使用的支持对象项目，并在预算范围内予以支持。支持的规模和程序可以通过科学技术信息通信部公告的各项指南进行确认。

4. 人工智能集聚园区的指定条件和入驻福利

法律第23条设立了人工智能集聚园区制度。施行令第17条规定了指定程序和考虑事项。

概念是这样的。它是人工智能企业、研究所和大学聚集在一处共享基础设施的空间。可以将其视为板桥科技谷或大田大德研究园区的人工智能特化版。

指定并非通过确定性的条件审查，而是通过考虑事项评估来进行的。根据施行令第17条第1款，科学技术信息通信部长官或地方自治团体的负责人综合考虑公告事项进行指定，包括是否符合基本计划、人工智能产业的区域集聚化及强化竞争力的效果、创造就业等对地区经济发展的贡献可能性、与其他集聚园区的重复性、关联性、可达性和效率性、专业人力的确保及可持续发展可能性等。指定事实以及位置、范围、区域将在相关机构网站上公告。

支持入驻企业的法律框架是法律第23条第2款。国家和地方自治团体可以对指定的集聚园区提供行政、财政和技术支持。减免租金或提供计算资源等具体福利并非法律保障的，而是由各园区的支持设计所决定的内容。規制特例不能依据本

法，而需经过规制沙盒等单独法制的程序才能获得。

截至2026年7月，光州正在运营人工智能中心产业融合集聚园区项目。对于首都圈以外的经营者来说，入驻集聚园区是解决计算基础设施成本问题的现实途径。

5. 人工智能研究所，2026年修订设立的新机构

2026年1月20日的修订法（法律第21311号）新设了第22条之2。即人工智能研究所。此次修订还同时新设了第22条之3，为科学技术信息通信部长官为了获取创新型人工智能技术而直接设立和运营研究机构开辟了道路。第22条之2自2026年1月22日起施行，第22条之3自2026年7月21日起施行。

人工智能研究所并非政府设立的机构。它是大学、企业、政府出资研究机构、非营利法人等单独或共同设立的民间研究所。由3名以上发起人组成并获得科学技术信息通信部长官的许可而设立，除本法规定的内容外，准用《民法》关于财团法人的规定。政府和地方自治团体可以在预算范围内支持运营经费，研究所可以接受补助金、捐赠金，或根据章程从事营利事业。

研究所执行的法定项目有五项（第22条之2第5款）。人工智能技术研发，与其他技术、跨学科融合的研发，研发成果的管理、转让、使用 and 商业化，专业人才培养，以及国际交流和国际合作。

对经营者而言，有意义的是成果的转让和商业化被纳入了法定项目。即使是没有资源从头开始创建大规模语言模型的中小企业，也开辟了获得研究所研发成果转让并将其应用于自身服务的途径。但具体的使用条件需通过与个别研究所的合同来确定。

设立的具体条件和程序已在施行令中制定。2026年7月21日施行的修订施行令第16条之2规定了八种设立资格者类型以及财务能力和安保对策条件，第16条之3规定了设立许可的申请程序。剩下的就是规定条件细节的科学技术信息通信部公告。

6. 风险投资母基金与促进人工智能投资

修订法在第18条新设的第3款至第5款是为了促进人工智能领域风险投资的条款。自2026年7月21日起施行。

母基金（fund of funds）是由政府出资和民间出资合并而成的大型基金。该母基金向人工智能专业风险投资（VC）基金出资，VC基金再投资个别人工智能企业，这是一个两阶段的结构。

以前也有中小风险企业部下属的韩国风险投资（KVIC）运营的母基金。然而，并没有专门用于人工智能的依据。第18条第3款使中央行政机关的负责人能够与中小风险企业部长官协商，利用《关于促进风险投资的法律》第70条规定的风险投资母基金提供支持。第4款规定其财源为国家、地方自治团体、公共机构的出资，希望出资的民间出资，以及附带收入。根据施行令第15条之3，中央行政机关的负责人可以要求韩国风险投资制定利用母基金的人工智能产业相关投资计划。这并非建立新的单独财源结构，而是将现有的母基金运用于人工智能领域的结构。

站在人工智能初创公司的立场，这是吸引投资的新途径。如果获得母基金出资的人工智能专业VC增加，种子轮和A轮阶段的投资规模和频率将会发生改变。根据中小风险企业部的统计，2025年韩国人工智能领域的风险投资规模约为1.3万亿韩元。政府的追加出资可以扩大这一规模。

经营者有一些需要了解的实务要点。母基金的出资条件会附加投资领域和企业经营年限限制。在人工智能专业VC基金的审查中，通常会要求提交证明经营者核心技术是人工智能的文件（专利、技术审查报告、政府项目执行履历）。这虽然不是法定条件，但在审查实务中是通用的标准。

7. 欧盟《数据治理法》与中国《数据安全法》，比较的实际利益

并非只有韩国为人工智能学习数据确立了法律依据。

欧盟的《数据治理法》（Data Governance Act）于2022年制定并生效，自2023年9月24日起适用。这是一部规定公共部门数据再利用条件的法律。虽然没有将在人工智能学习中作为明确目的提及，但它开辟了在安全处理环境（secure processing environment）中出于研究和创新目的再利用非识别数据的道路。创造了数据中介服务（data intermediation service）这一新的业务类型，该中介者在数据持有者和使用者之间调解可信的交易。

韩国的综合提供系统是政府直接运营的集中型。欧盟则是建立由通知和注册体系管理的民间中介者的分散型。两者理念不同。

中国的《数据安全法》（2021年施行）实行数据分类分级保护制度。国家核心数据实行更严格的管理，重要数据实行目录管理，并根据级别适用不同的保护水平。可用于人工智能学习的数据范围会根据这种定级分类而缩小或扩大。由于重要数据的出境会附加安全评估等单独程序，在韩国运营人工智能服务的企业在中国通常会选择在当地处理学习。

《生成式人工智能服务管理暂行办法》（2023年施行）要求使用具有合法来源的数据和基础模型，不得侵害他人的知识产权。如果有监督机构的检查，必须说明学习数据的来源、规模、类型等。韩国法律中尚未规定这种程度的数据来源说明义务。

如果韩国经营者同时进入这两个市场，从一开始就设计成满足三个法域要求的数据治理文件是高效的。

8. 知识产权保护与人工智能学习的紧张关系

《人工智能基本法》中没有将知识产权保护设为独立条文的规定。最接近的条款是第13条第3款。该规定指出，当政府公开和普及旨在实现安全使用人工智能的支持项目成果时，为了保护技术开发者，可以设定保护期限使其收取技术使用费，或通过其他方法予以保护。人工智能技术本身的知识产权（专利、商业秘密）和学习数据的权利，受《专利法》、《防止不正当竞争法》、《著作权法》等一

般知识产权法制的规制。

这与第14章讨论的著作权争议角度不同。著作权是为了防止未经授权使用创作物，而这里涉及的是人工智能技术本身及其成果的保护。

经营者在政府支持下开发的人工智能模型的知识产权归属于谁？这个问题尚不明确。政府研发项目成果的知识产权归属遵循《产业技术创新促进法》和《国家研发创新法》，但在人工智能领域，模型权重（model weights）、学习数据、学习方法论的权利归属可能会各有不同。

经营者对人工智能研究所（第22条之2）开发的基础模型进行微调而生成的衍生模型的知识产权如何处理？关于这个问题的施行令层面的指南尚未出台。经营者在与研究所签订技术转让合同时，必须在合同中明确规定衍生模型的权利归属。

9. 经营者为获得政府支持需要准备的事项

法律建立的支持基础设施主要有四个。学习用数据综合提供系统（第15条）、人工智能集聚园区（第23条）、利用风险投资母基金的投资财源（第18条第3款）、人工智能研究所的研究成果（第22条之2）。要接触这四个方面，经营者这边也需要做好准备。

申请使用数据需要使用目的、安保对策和个人信息处理计划。必须具体说明获取数据后要制作什么，以及为了防止数据泄露到外部将采取哪些技术措施。

如果需要数据质量认证，必须确认的不是本法，而是《数据产业法》体系的质量认证制度。《人工智能基本法》只是将通过综合提供系统提供的数据的质量保持责任交给了政府，并没有设立针对经营者数据集的认证制度。

入驻集聚园区遵循公开招募程序。入驻申请书需填写业务计划、技术人力现状和预期成果。现有集聚园区的入驻竞争率和审查标准可在专门机构（根据法律第23条第4款由政府设立或指定的机构）或相关地方自治团体的公告中确认。

吸引投资的方式是在韩国风险投资（KVIC）网站上确认获得母基金出资的人工智能专业VC名单，并向相关VC提交商业计划书（IR deck）。如果附上证明是否拥有人工智能核心技术的材料（专利注册证、政府项目执行确认书、技术保证基金评估书），将有利于审查。

政府的支持不是免费的。作为获得支持的代价，伴随着成果报告、事后管理和提交数据使用业绩的义务。如果不履行这些义务，将成为收回支持金的事由。在申请支持之前，必须将事后管理的负担也计算在内。

如果是在韩国国内获得政府支持同时进军欧盟市场的经营者，必须确认政府支持数据是否可以转移到海外。从综合提供系统获取的部分数据可能会附加仅限国内使用的条件。如果欧盟法人需要直接访问相关数据，需审查数据转移合同，如果包含个人信息，则必须具备《个人信息保护法》规定的向境外转移的条件之一，如单独同意、履行合同所需的处理委托及保管、充分性决定等。

第18章. 公共采购的优先考虑与公务员免责条款 —— 开启B2G市场的钥匙

2025年秋，首尔某区厅福利政策科的负责人未能回答区厅长的问题：“用AI抓取违规领取补助的系统，据说每年能节省数十亿韩元，为什么不引入呢？”负责人的内心想法是这样的：如果系统发生故障，错误地淘汰了合法的领取者，做出采购决定的自己在监查院的监查中可能会承担赔偿责任。企业方面要求在合同中加入免责条款，但法制负责人以没有先例为由面露难色。既有技术又有预算，但公务员个人的赔偿责任却阻碍了引入。2026年1月20日公布的韩国《人工智能基本法》修订法为了打破这一僵局，在第16条中增加了两款。

1. 国家机关应优先考虑AI

修订法第16条第3款这样写道：

“国家机关等在采购执行业务所需的产品或服务，或者发包劳务时，应当优先考虑总统令规定的人工智能产品或人工智能服务。但是，根据业务特性不适合运用人工智能技术的情形除外。”

句子虽短，但具有影响力。“国家机关等”包括国家、地方自治团体以及根据《公共机关运营相关法律》第4条规定的公共机关（第2条第4项）。“应当”是义务规定。不是裁量。这意味着采购负责人在购买产品或服务时，必须首先探讨是否存在AI替代方案。

但书条款起到了缓冲装置的作用。在“根据业务特性不适合运用人工智能技术的情形”下，免除该义务。这并不是说在购买铅笔和复印纸时也要考虑AI。

法律设定了“总统令规定的人工智能产品或人工智能服务”这一限制。《施行令》第15条第4款将其规定为“科学技术信息通信部长官确认的产品或服务”。并不是任何AI产品都能成为优先考虑的对象。必须经过科学技术信息通信部长官确认这一关卡。确认的标准和程序由科学技术信息通信部长官与行政安全部长官

协商后公告（《施行令》第15条第5款）。

2026年5月21日进行立法预告的《施行令》修订案具体化了这一确认程序。其结构是，韩国人工智能振兴协会确认相关产品是否应用了AI技术，韩国信息通信技术协会（TTA）进行技术审查。具体程序预计将根据科学技术信息通信部长官确定并公告的标准来制定。

从企业立场来解释这一结构如下。如果要向公共市场供应AI产品，必须经过韩国人工智能振兴协会和TTA的审查，并获得科学技术信息通信部长官的确认。获得确认的产品在国家机关的采购过程中，将比非AI替代方案优先得到探讨。

韩国以往也存在类似的制度。《采购事业相关法律》第27条的创新产品公共采购支持制度（2020年施行）、采购厅的创新产品示范采购项目、创新市场（创新产品公共采购综合信息系统）、企划财政部训令《关于创新产品指定及促进采购等规定》等便是。现有的创新产品制度赋予了采购厅长“可以支持”创新产品公共采购的裁量权。韩国《人工智能基本法》第16条第3款赋予了“应当优先考虑”的义务。这是从裁量转变为义务。

2. 称为免责的安全装置

我们来看第16条第4款。

“因采购或使用第3款规定的人工智能产品或人工智能服务而对国家机关等造成损害的，负责该采购或使用业务的人员对该机关的损失不承担赔偿责任。但是，因该业务负责人的故意或重大过失造成损失的，不在此限。”

要理解为什么需要这一条款，必须考虑公务员的处境。AI是概率性运作的。结果取决于训练数据，在引入时无法进行完美验证。如果是传统软件，对于输入值会产生相同的输出值。AI则不同。对于相同的输入，可能会根据概率分布产生不同的结果。

公务员引入AI系统后，如果因故障给机构造成损失会怎样？根据《国家财政法》，预算执行的适当性可能会成为问题。如果在监查院的监查中出现“在没有充分探讨的情况下就引入”的指责，采购审批流程上的负责人就会成为被追责的对象。甚至可能面临根据《公务员惩戒令》的纪律处分以及根据《国家赔偿法》的求偿权行使。

理性的公务员会怎么做呢？维持现有系统。不引入新技术对个人来说是安全的。行政学称之为“风险规避偏见”（risk-aversion bias）。这是公共部门创新缓慢的一个常见原因。什么都不做就不需要承担责任。尝试了什么却失败了，就要承担责任。在这种不对称的结构下，创新就停止了。

第16条第4款通过法律纠正了这种不对称。购买或使用经科学技术信息通信部长官确认的AI产品的公务员，即使因此给机构造成损害，也不承担赔偿责任。

免责的条件有两个。第一，购买的产品必须是第3款规定的优先考虑对象。意思是获得了科学技术信息通信部长官确认的产品。如果擅自购买未获确认的AI产品，就不属于免责对象。第二，不得有故意或重大过失。

我们具体划定界限。

引入了AI招聘评估系统，却得出了对特定性别不利的结果。制造商建议“每季度进行偏见检查”，但负责人在2年内一次也没有检查。这可能构成重大过失。

相反，如果引入相同的系统，并按照制造商手册每季度进行检查，但因无法预测的数据异常而发生错误，就可以免责。

有着类似结构的先例。《采购事业相关法律》第27条第4款规定，购买创新产品的需求机关的采购负责人，只要未被证明存在故意或重大过失，不对因购买该产品造成的损失承担责任。免除采购负责人的责任但将故意和重大过失列为例外的结构是相同的。韩国《人工智能基本法》第16条第4款立足于这一谱系之上，是正面处理AI概率性不确定性这一新问题的条款。

这对企业具有重要意义。公共客户不愿采购人工智能的巨大心理障碍已被法律降低。在销售现场，企业可以解释说：“只要没有故意或重大过失，负责人您就不必对机构损失进行赔偿，法律中有免责条款。”

3. 影响评估开启的第二扇门

第16条的优先考虑义务适用于科学技术信息通信部长官确认的人工智能产品和服务。对于高影响人工智能，还有两个单独的优先考虑机制。

即第35条第2款：“国家机关等拟使用利用了高影响人工智能的产品或服务的，应当优先考虑已实施影响评估的产品或服务。”

影响评估本身并非强制性规定。第35条第1款规定了“应当努力”。这种“努力义务”是不带制裁的义务。不履行也不会被处以行政罚款。事实上，查看第43条行政罚款条款可知，被处以行政罚款的对象只有违反透明度告知义务（第31条第1款）、未指定韩国境内代理人（第36条第1款）、不履行停止命令或纠正命令（第40条第3款）这三种情况。没有因为不进行影响评估而处以行政罚款的条款。

然而第2款为这种自主性加上了市场的重量。完成影响评估的产品在公共采购中会被优先考虑。未做的产品则被推后。

《施行令》第28条列举了影响评估中应当包含的7项内容：可能受到影响的个人或群体的识别（包括反映弱势群体的特征）、基本权利类型的识别、社会和经济影响的内容与范围、使用形态、定量或定性评估指标与结果计算方式、风险的预防与缓解及损失恢复方案、改善实施计划。经营者可以直接执行，也可以委托第三方进行（《施行令》第28条第2款）。

第30条第4款也使用了同样的逻辑：“国家机关等拟使用高影响人工智能的，应当优先考虑基于已获得检验和认证等人工智能的产品或服务。”接受了安全性与可靠性自主检验或认证的AI产品，在公共采购中会进一步被优先考虑。

将这三个条款并列来看，其结构就显而易见了。

第16条第3款规定优先考虑科学技术信息通信部长官确认的AI产品。

第30条第4款规定优先考虑获得检验和认证的高影响人工智能产品。

第35条第2款规定优先考虑实施了影响评估的高影响人工智能产品。

是三层。同时具备科学技术信息通信部确认、检验和认证、影响评估的产品，将同时通过三个关卡。一样都不具备的产品，在公共市场上的竞争条件将变得不利。

4. 软监管的逻辑

这就是韩国《人工智能基本法》的“软监管”结构。

法律并未直接强制要求进行影响评估或认证。不征收行政罚款。取而代之的是，以政府采购市场这一大规模需求为媒介，引导企业自愿遵守。不守法没有惩罚，但守法就能获得市场。

韩国法律的行政罚款上限为3000万韩元（第43条）。欧盟《人工智能法案》（EU AI Act）的罚款上限根据违规类型而异。违反被禁止的人工智能惯例（Article 5），最高可处以3500万欧元（约500亿韩元）或全球营业额7%的罚款；在未经合格评定的情况下推出高风险人工智能等违反经营者义务的行为，最高可处以1500万欧元或全球营业额3%的罚款。在韩国的3000万韩元和欧盟的1500万欧元之间，也有约800倍的差距。

韩国立法者是故意制造这种差距的。在判断国内AI产业正处于初期成长阶段的基础上，接受了企业方面过度的制裁可能会萎缩创新的逻辑。大棒虽小，但胡萝卜很大。3000万韩元的行政罚款很难给大企业提供遵守的动力，但可能失去整个政府采购市场的前景则成为遵守的动力。

这种方法并非韩国独有。将政府采购作为遵守监管的诱因的做法，在美国也出现过。2023年拜登政府的AI行政命令（Executive Order 14110）规定，联邦机

构在采购AI时优先考虑满足安全性和透明度要求的产品（该行政命令于2025年1月被废除）。韩国《人工智能基本法》第16条在法律层面规定了类似的逻辑。

区别在于，美国的是行政命令，总统换人就可能被废除。2025年1月，特朗普政府撤销了拜登的AI行政命令。韩国的是法律。只要国会不修改，它就会继续存在。

5. B2G市场的规模

韩国公共部门对AI的需求可以从制度指标中看出。

48个中央行政机关已全员指定了首席人工智能官（CAIO）。部长级机关指定副部长级，厅等副部长级组织指定室长级。自2025年9月18日起，每月举行定期协商会。国家人工智能战略委员会由总统担任委员长，委员总数不超过60人，下设8个分科委员会，有52名分科委员，共计85人的体制。

2026年2月25日，在战略委员会第二次全体会议上，《大韩民国人工智能行动计划（2026-2028）》被确定。包含99项执行课题，326项政策建议。48个机构的负责人将在三年内执行99项课题。其中相当一部分课题很可能会导致人工智能产品和服务的采购。

人工智能政策中心于2026年4月29日在韩国智能信息社会振兴院（NIA）内成立，并于6月30日正式揭牌。下设3个队、1个团共5个队，拥有59名人员。支持制定基本计划、支持政策开发、调查AI扩散影响是其法定业务。人工智能安全研究所（AISI）作为韩国电子通信研究院（ETRI）的附属机构独立运营，负责安全评估和安全技术研究。政策规划机构、安全研究机构、经营者支持机构均已齐备。剩下的就是实际采购了。

第16条第3款的优先考虑义务自2026年7月21日起施行。自此日期后，中央行政机关、地方自治团体和公共机关的采购负责人在采购产品或服务、或发包劳务时，必须首先探讨AI替代方案。韩国的国家机关、地方自治团体、公共机关加起来

有数千家。各机构的信息化业务、行政服务和直接面向市民的业务中对AI转型的需求，由法律创造出来。

韩国在联合国电子政务发展指数（EGDI）中，连续多年名列前茅。“政府24”、“国民秘书”、国税厅“Hometax”等数字化公共服务基础设施已经铺设完毕。在这些基础设施上加入人工智能，在技术上是顺理成章的下一步。第16条第3款为这下一步赋予了法律义务。

6. 企业应准备的事项

向公共部门供应人工智能产品的企业应当确认的事项，按顺序整理如下。

首先，判断自家产品是否属于高影响人工智能。法律第2条第4项列举了11个领域（能源供应、医疗保健、医疗器械、核能安全、生物特征信息分析、招聘或贷款审查等对权利产生重大影响的判断、交通管理、公共机关的资格审查和收缴服务费、学生评价等）。如果属于这些领域，就是高影响人工智能。如果不清楚是否属于，可以向科学技术信息通信部长官请求确认（第33条第1款）。

如果是不属于高影响人工智能的一般人工智能产品，只要获得科学技术信息通信部长官的确认（《施行令》第15条第4款），就会进入优先考虑对象之列。这是经过韩国人工智能振兴协会的技术应用确认和TTA的技术审查的程序。

如果是属于高影响人工智能的产品，满足所有三项优先考虑条款将是有利的。

科学技术信息通信部的确认（与第16条第3款关联的《施行令》第15条第4款）是基础。

如果实施了影响评估（第35条），就会被纳入第二个优先考虑对象。必须撰写包含《施行令》第28条规定的7项内容的评估报告。包括识别可能受到影响的个人或群体（反映弱势群体的特征）、识别基本权利类型、分析社会和经济影响的范围、描述使用形态、提出评估指标与结果计算方式、制定风险的预防与缓解及损失恢复方案、加入改善实施计划。可以自己执行，也可以委托第三方专业机构

执行。

如果获得了检验和认证（第30条），就会被纳入第三个优先考虑对象。这是由科学技术信息通信部长官支持的自主安全性与可靠性检验或认证。

如果三者兼备，就满足了所有的三个优先考虑条款。在公共采购中的竞争优势将变得明确。

还有一项需要确认的，那就是人工智能基本法支持台。由韩国软件产业协会（KOSA）运营，截至施行第10周的2026年3月31日，共收到552件咨询。可以进行匿名咨询，保护商业机密。可以就属于高影响人工智能与否或影响评估方法进行咨询。

7. 欧盟选择了沙盒

欧盟《人工智能法案》（EU AI Act）推出了监管沙盒（regulatory sandbox）机制，而不是公共采购优先考虑。第57条至第62条（Article 57~62）规定了该制度。

监管沙盒是指在经营者将产品实际投放市场之前，在监管机构的监督下能够开发和测试人工智能系统的受控环境。欧盟《人工智能法案》规定每个成员国必须设立至少一个人工智能监管沙盒。原定期限是2026年8月2日，但由于数字法案集（Digital Omnibus）的修订，推迟一年至2027年8月2日。

西班牙在2022年6月法案确定之前就启动了示范沙盒，这是欧盟首例。然而，截至2026年7月，大多数成员国尚未正式开设人工智能监管沙盒。将期限推迟一年本身就暴露了准备不足。

在沙盒内，企业在监管机构的监督和建议下，在限制条件下测试高风险人工智能系统。针对中小企业和初创企业有减免手续费等优惠措施。也允许在实际环境中进行有限测试（real-world testing），这种情况下需要使用者的事先同意（Article 60~61）。

韩国和欧盟的方法不同。

欧盟采取的是事前监管的方法。在将高风险人工智能投放市场之前，必须进行合格评定（conformity assessment）。监管沙盒是为这个合格评定做准备的实验室。先有监管，沙盒是用于通过监管的工具。

韩国采取的是市场诱导的方法。法律不强制要求影响评估和认证。取而代之的是，以公共采购这个市场为媒介，引导企业自愿遵守。监管在后，市场在前。

韩国也有实证基础。法律第24条规定了“造成人工智能实证基础”，允许国家和地方自治团体建立和运营用于支持实证和性能测试的设施、装备和设备。与欧盟的监管沙盒外表相似，但目的不同。欧盟的沙盒是为了帮助遵守监管的环境。韩国的实证基础是为了培育产业的基础设施。

也有对欧盟沙盒的批评逻辑。在沙盒内通过的人工智能无法保证在实际市场复杂环境中依然安全。在受控环境下的成功并不意味着在非受控环境下的成功。另一方面，也有对韩国软监管的批评。对于无意进入公共采购市场的企业，根本没有动机进行影响评估。仅面向民间市场的人工智能产品，享受不到软监管的胡萝卜。韩国法律中3000万韩元的行政罚款，对于要在民间领域强制遵守来说，金额简直微不足道。

8. 门打开了

2026年7月21日以后，韩国公共部门将负有优先考虑获确认的AI产品的法定义务。公务员受到免责保护。企业必须具备科学技术信息通信部确认的资格。

首尔那个区厅福利政策科的负责人现在有了法律上的答案。可以引入人工智能违规领取补助检测系统，即使因故障造成损失，只要按照手册操作，就不必对机构的损失承担赔偿责任。有个条件：必须是科学技术信息通信部长官确认的产品。

法律敞开了大门。谁能通过那扇门，取决于《施行令》公告的确认标准、韩国人工智能振兴协会和TTA的审查能力，以及企业推出产品的水平。公共市场比民

间市场慢，程序长，需要熟悉行政语言。但作为回报，一旦进入，合同将以数年为单位，而且在机构间的扩散速度很快。在一个部委取得成功的AI服务，可以通过48个机构的CAIO协商会传播到其他部委。

法律创造的是市场之门。市场本身则是由产品创造的。

第19章。弱势群体无障碍支持与费用支持，2026年修订开启的新市场

让我们想象一下住在全南某农村的78岁金奶奶试图使用农业用AI病虫害诊断应用的场景。虽然打开了孙女帮她安装的应用，但字号太小，且语音识别无法准确听懂当地的方言。尝试了三次后她放弃了。无论AI技术多么精妙，对于无法接触到该技术的人来说，它就如同不存在一样。这就是2026年修订法新增第17条之2、开辟使用费用支持途径的原因。

1. 修订前不存在的条款，修订新增的第17条之2

原法律（2025年1月21日公布，2026年1月22日施行）中并没有关于弱势群体AI无障碍获取或费用支持的独立条款。在基本计划中包含确保无障碍获取的内容，差不多就是全部了。

2026年1月20日的修订法（法律第21311号，2026年7月21日施行）填补了这一空白。第3条第5款将使用人工智能产品或人工智能服务面临困难的残疾人、老年人等称为“人工智能弱势群体”，并规定国家和地方自治团体有义务保障其参与人工智能政策的开发与制定过程，并努力使其意见得到反映。第6条第2款第7项规定，人工智能基本计划必须包含保障弱势群体接触和使用的事项。而第17条之2则为因经济条件难以使用AI产品与服务的人群，提供了在预算范围内支持全部或部分费用的新依据。

弱势群体的范围由《施行令》第1条之2规定了十种类型：根据《求职者就业促进法》具备受领资格的人、基础生活保障受领权者和次上位阶层、农渔业从业者、多文化家庭成员、北韩脱离居民、保有工作履历的女性、残疾人、单亲家庭支持对象者、65岁以上的人，以及经中央行政机关或地方自治团体的首长认定在使用上面临困难的人。费用支持对象由《施行令》第15条之2另行规定，除上述类型外，还包括理工科人才、被评定伤残等级的国家有功者以及区域均衡人才。

过去仅停留在宣言层面的无障碍获取，如今转变成了费用支持的依据。这一变化对经营者而言意义明确。政府可以通过编制预算来补贴特定群体使用AI服务的费用。从经营者的立场来看，这意味着出现了一个有政府补贴流入的新需求群体。

2. 费用支持开启的市场规模

我们用数据来看。登记在册的残疾人约有263万人，65岁以上的老年人约有1050万人。截至2024年底，基础生活保障受领者约有267万人；作为保有工作经历女性的核心群体，因生育等中断职业生涯的女性（职业中断女性）截至2025年约为110万人；农渔业从业者约为230万人。再加上多文化家庭成员、北韩脱离居民、单亲家庭支持对象者、次上位阶层，而作为费用支持对象，理工科人才、伤残等级国家有功者、区域均衡人才也将纳入其中。

关于求职者，阅读时需要注意。《施行令》规定的对象并非一般的求职者，而是根据《求职者就业促进及生活稳定支援相关法律》具备受领资格的人，即在国民就业支持制度中被认定具备受领资格的人。由于这与雇佣保险失业给付受领者是不同的制度，因此不能将这两个群体的规模混在一起计算。

即使排除重复计算，潜在受益人口也达数百万规模。假设政府向每人每月发放价值1万韩元的AI服务使用券，每年将创造出数千亿韩元规模的市场。但这终究只是基于假设的计算。法律仅规定了可以给予支持的依据，对于是否引入代金券、具体金额、受益规模等尚未做任何规定。

实际规模将取决于预算编制。在2026年下半年施行后不久，很有可能会以试点项目的形式分配小规模预算。但政策方向是倾向于扩大。一旦支持使用AI服务的政策正式铺开，对于教育、医疗、农业领域的人工智能经营者而言，这将成为稳定的需求基础。

3. 残疾人AI无障碍获取，保障什么以及如何保障

《人工智能基本法》和《施行令》目前尚未另行制定残疾人无障碍获取的技术标准。在实际业务中通用的标准源自《禁止残疾人歧视法》及其《施行令》，以

及《韩国型网页内容无障碍指南》（KWCAAG）。如果将 these 标准应用于AI服务，就能看出需要准备的内容。

针对视障人士，需要提供与屏幕阅读器（screen reader）兼容的界面。如果AI聊天机器人以文本形式进行回复，该文本必须能够被TTS（从文本到语音）引擎读取。如果是生成图像的AI，则需要具备在生成的图像上添加替代文本（alt text）的功能。由于自动生成的替代文本准确性可能会有波动，因此同时设置人工审核流程会更加安全。

针对听障人士，需要在基于语音的AI服务中提供字幕或手语翻译功能。如果AI语音助手仅以语音形式提供回答，而没有提供文本转换选项，听障人士就无法使用该服务。

针对肢体残疾人，需要提供仅通过键盘即可操作的界面，以及通过语音命令进行控制的功能。如果鼠标点击或触摸屏操作是唯一的输入手段，那么该AI服务就无法满足无障碍标准。

这些要求并非新创，而是原有的《禁止残疾人歧视法》和网页无障碍标准在AI领域的延伸。然而，由于AI服务的特性，需要进行额外的考量。生成式AI所生成内容的无障碍访问责任由谁承担？如果使用经营者用AI制作的营销图像没有替代文本，这究竟是使用经营者违反了无障碍义务，还是生成式AI开发经营者的功能缺陷？这种责任界限目前在任何法令中都尚未明确界定。

4. 老年人AI普及，跨越数字鸿沟的设计

面向老年人的AI服务，其关键在于界面设计，而非技术本身。

根据科学技术信息通信部和韩国智能信息社会振兴院进行的《数字信息差距状况调查》，70岁以上老年人的智能手机拥有率已达到90%左右。然而，以普通国民的数字素养为100分进行衡量，老年人的利用水平仅停留在40多分。虽然拥有设备，但无法充分使用。

AI服务经营者若想进入老年人市场，大号字、高对比度、语音优先的界面是基础。此外，还需要有容易撤销失误的设计（明确布置undo功能）、不含专业术语的引导语，以及拨打电话即可连通的人工客服作为支撑。

《施行令》第1条之2第9项将65岁以上人群纳入人工智能弱势群体，这释放了一个信号，即政府将向进行此类设计的经营者提供对接支持。开发面向老年人定制化AI服务的经营者，值得关注科学技术信息通信部的数字包容相关项目公告，并确认参与条件。

5. 保有工作履历的女性与求职者，AI教育与服务支持

将保有工作履历的女性和具备受领资格的求职者纳入弱势群体是较为罕见的。因为这不是根据残疾或年龄等身体条件，而是根据经济、社会状况进行的分类。法令所使用的名称也值得关注。不是通常所说的“职业中断女性”，而是《促进女性经济活动及预防职业中断法》中的“保有工作履历的女性等”，且求职者被限定为国民就业支持制度中的受领资格者。

立法宗旨是为了应对由AI引起的就业结构变化。如果AI替代现有岗位的速度快于能够使用AI的劳动力的转型速度，就会有人在这一差距中被淘汰。保有工作履历的女性和求职者是极易处于这种差距之中的群体。

对于经营者而言，这意味着AI教育服务市场的扩大。如果政府向他们发放AI应用教育代金券，就需要有接受该代金券并提供教育培训的经营者。提示词工程培训、AI辅助工具使用方法培训、分行业AI应用实操培训将成为新的需求。

如果能与性别平等家族部的保有工作履历女性支持项目、雇佣劳动部的“国民明日学习卡”制度相对接，预算规模可能会扩大。但这种对接并非法律规定的必然效果，而是取决于政策设计的可能性。经营者需要确认自己的AI培训程序是否能被纳入这一政府支持体系中。认证条件将遵循相关部门的教育机构注册标准。

6. 农渔业从业者与区域均衡人才，一线的AI普及

将农渔业从业者纳入对象（《施行令》第1条之2第3项），是反映了缩小地区差距这一政策目标的结果。

非首都圈的中小企业即使有引入AI的意愿，也缺乏人才和基础设施。如果委托给首尔的AI咨询公司，光是出差费就达数百万韩元。虽然云GPU费用与首都圈相同，但在周边找不到能在现场清洗数据并应用模型的人才。然而，法令中并没有名为“非首都圈中小企业在职人员”的支持范畴。《施行令》规定的是指地方大学出身人才的“区域均衡人才”（第15条之2第4项），以及经中央行政机关或地方自治团体的首长认定在使用上面临困难的人（第1条之2第10项、第15条之2第5项）。虽然地方自治团体可以将地区产业人才纳入该认定条款中，但这必须等公告发布后才能确定。

农渔业从业者的情况更为严峻。虽然智能农场、病虫害诊断、产量预测、养殖场水质管理等AI应用领域非常明确，但一线的数字基础设施和用户能力却跟不上。

政府的费用支持缩小了这一差距。第17条之2直接规定的是支持使用费用。针对现场的定制化培训或设备普及，则需要有单独的项目依据和预算支持。不过，像地方自治团体的智能农场支持项目那样，将多种支持捆绑在同一个公告中发布的情形完全是可以预见的。对于智能农场AI经营者、农业无人机分析经营者、养殖管理AI经营者而言，将产生一个有政府补贴的稳定客户群。

经营者要想进入这一市场，需要做几项准备。首先，追踪农村振兴厅、海洋水产部、中小风险企业部的相关支持项目公告。其次，申请作为AI服务供应商参与非首都圈地方自治团体的智能农场与智能养殖支持项目。此外，还需配备离线模式或低带宽环境应对功能，以便在现场使用。

7. 《欧盟无障碍法》与中国《无障碍环境建设法》的对比

《欧盟无障碍法》（European Accessibility Act, Directive (EU) 2019/882）已于2025年6月28日起适用。该法并仅不针对AI。它对电脑、智能手机、ATM、电子商务网站、交通服务等列出的产品和服务的无障碍获取进行了规制。如果AI服

务与这些列出的产品或服务相结合，也将一同适用该法，因此进军欧盟市场的韩国AI经营者也必须在此范围内满足无障碍获取的要求。

在《欧盟人工智能法案》中，高风险AI的无障碍获取由第16条规定。高风险AI系统的提供者有义务对系统进行设计，以满足《公共部门网页无障碍指令》（Directive (EU) 2016/2102）和《欧洲无障碍法》规定的无障碍要求（第16条第(l)项）。

韩国法律与欧盟法律的区别在于方式。韩国是在需求侧给予支持，为弱势群体补贴费用。而欧盟则是在供给侧进行规制，强制要求经营者满足无障碍标准。如果韩国经营者进军欧盟市场，就会面临双重结构：在国内是费用补贴所创造的需求的受益者，而在欧盟则是履行无障碍业务的合规义务主体。

中国的《无障碍环境建设法》（2023年9月施行）保障残疾人和老年人的信息无障碍获取。该法要求对网站和移动应用进行无障碍改造，虽然没有单独明确提及AI服务，但只要AI是通过网页和应用提供的，就很有可能被解释为适用相同的要求。惩戒以纠正命令和通报批评为主，罚款仅限于电信业务经营者等部分主体的违规行为。

在这三个国家中，韩国的第17条之2在对象范围上非常独特。欧盟和中国的无障碍法律主要集中于残疾人和老年人，而韩国则涵盖了保有工作履历的女性、具备受领资格的求职者、农渔业从业者以及北韩脱离居民等社会经济弱势群体。这一范围的扩大能否得到实际预算和政策的支持，是法律施行后的重要看点。

8. AI无障碍获取市场的规模与准入策略

目前很难估算准确的市场规模。因为政府预算尚未确定，且支持对象和标准是由中央行政机关的首长和地方自治团体的首长根据《施行令》第15条之2以公告形式规定的。公告中将包含考量申请人使用AI的必要性，经济、地区、身体、社会状况，以及预算规模和支持公平性等因素的标准。

但可以读出其发展方向。科学技术信息通信部每年都在运营数字包容相关项目。如果将支持AI使用费用的内容纳入该项目体系，初期可能会以试点项目的规模起步，在确认成效后有扩大的可能。

经营者的准入策略因目标群体的不同而有所差异。面向残疾人与老年人，辅助器具联动、语音界面、自动生成替代文本等技术能力是核心。面向保有工作履历的女性与求职者，培训内容的质量以及结业后与就业对接的成效是评估标准。面向农渔业从业者，现场适用性（应对网络不稳定环境、兼容防水防尘设备）则具有决定性意义。

有几项需要共同准备的事项。首先，提前掌握政府支持项目的招标与公告流程。其次，获取残疾人无障碍认证（网页无障碍质量认证标志）。虽然这并非法定强制认证，但在公共和福利领域的招标中，可以作为证明自身是做好准备的经营者的依据。最后，建立使用者满意度评估体系。由于政府支持项目通常会根据使用次数、满意度、二次使用率等绩效指标进行评估，因此必须在服务中内置收集这些指标的系统。

9. 与现有无障碍法律的关系

第17条之2并非独立运作。它与《禁止残疾人歧视法》、《残疾人福利法》、《国民基础生活保障法》、《多文化家庭支持法》、《关于北韩脱离居民保护及定居支持的法律》、《单亲家庭支持法》、《促进女性经济活动及预防职业中断法》、《农渔业从业者生活质量提高特别法》等现有法律相挂钩。

《禁止残疾人歧视法》已经禁止了信息获取方面的歧视。如果提供的AI服务无法供残疾人使用，根据服务类型和提供合理便利义务的范围，可能会产生违反该法的问题。第17条之2并非禁止性规范，而是在此基础上增加了积极支持（补贴使用费用）的条款。

虽然法律并没有规定这两种制度的先后顺序，但实际业务的先后顺序是显而易见的。在因无障碍设施不足而收到违反《禁止残疾人歧视法》的纠正建议的情况

下，去申请政府支持项目是缺乏说服力的。首先满足禁止歧视这一法定合规义务，在此基础上抓住政府支持（费用补贴、项目参与）的机会，才是稳妥的途径。

《农渔业从业者生活质量提高特别法》将扩充农渔村的信息通信基础设施列为公共服务项目。为了使第17条之2的费用支持能够惠及农渔业从业者，农林畜产食品部与科学技术信息通信部之间的预算与项目协调将是实际业务中的变量。

对经营者而言，弱势群体的无障碍获取并不是负担，而是机会。因为这种结构是由政府创造需求、补贴费用，并交由经营者进行供给。这一结构能否正常运转，将取决于各部委和地方自治团体的公告以及预算编制。

第20章 创业·专业人才·创业投资母基金·人工智能研究所 —— 获取资金与人才的方法

2025年秋天，在首尔江南的一家共享办公室里，三位AI初创公司的代表一起吃了午饭。其中一人正在开发医疗影像解读AI，一人正在研发制造工序优化模型，还有一人正在准备教育用生成式人工智能。三人的共同烦恼并非技术，而是资金和人才。种子投资资金已见底，却未能找到下一轮投资者，而要招聘硕士级别的AI工程师，还必须与大企业竞争薪资。虽然他们知道《人工智能基本法》将于2026年1月22日起施行的消息，但脑海中只浮现出监管义务，并未想到该法能给他们带来些什么。

《人工智能基本法》并非只是监管法。在法律全部46个条款中，从第13条至第26条共17个条款被用于技术开发和产业培育。2026年1月20日的修改法在此基础上又增加了四项依据：创业投资母基金的利用、弱势群体使用费用的支持、提供公共数据作为学习用数据，以及政府直接设立研究机构。上述三位代表所寻找的关于资金和人才的法律依据，其实早已制定好了。

1. 创业支持：法律开辟的七条通道

《人工智能基本法》第18条以“创业的活跃化等”为题，列举了7项业务：创业者的发掘与培育、创业教育与培训、支持专业人才的优秀技术商业化、AI技术的价值评估与创业资金的金融支持、提供研究及技术开发成果、培育创业支持机构与团体，以及其他促进创业活跃化所需的业务。

这里有两个值得关注的部分。其一是第18条第1款第4项的“价值评估及创业资金的金融支持”。AI技术初创公司难以逾越的关卡是将技术价值用数字来证明。它们几乎没有有形资产，还没有销售额，拥有的只有模型性能指标和数据集。这相当于法律赋予了政府建立评估AI技术价值体系的依据。

其二是第18条第2款。其中规定，地方自治团体可以向支持AI创业的公共机构或公共团体进行捐助或出资。这意味着首尔市或京畿道可以设立AI创业专属的加速器机构并为其投入预算。如果将基层地方自治团体也包括在内，全国243个地方自治团体在任何地方都拥有了向AI创业支持机构投入公共资金的法律依据。

对比十分鲜明。欧盟《人工智能法案》（EU AI Act）是一部监管法。虽然有针对性对中小企业的照顾措施（如优先进入监管沙盒），但并没有直接支持创业资金或投资的条款。欧盟通过另一项政策包——“InvestAI”来推进AI产业的振兴。欧洲委员会于2025年2月公布的这一政策包，目标是实现政府与民间共计2,000亿欧元的投资。在法律中写入振兴条款的韩国，与将监管法与振兴政策相分离的欧盟，在准入路径上本身就截然不同。

中国则又有所不同。2017年国务院发布的《新一代人工智能发展规划》是AI产业振兴的根基。其出发点并非法律，而是国务院文件，结构上是由地方政府竞争性地提供产业补贴、土地和税收减免。深圳市向AI企业发放研究补贴，杭州市向AI人才提供住房补贴，这些并不是法律的命令，而是地方政府的裁量性产业政策。韩国《人工智能基本法》第18条在基本法层面确立了此类支持的法律依据，在体系上与此不同。

2. 创业投资母基金：政府资金流向初创公司的机制

第18条第3款是核心。“中央行政机关首长为了促进人工智能产业领域的创业活跃化，可以与中小风险企业部长官协商，利用创业投资母基金进行支持。”

创业投资母基金是依据《创业投资促进法》第70条设立的FOF（基金中的基金，fund of funds）。它由韩国风险投资股份有限公司负责运营。政府和公共机构向母基金出资，母基金再将这笔资金投入创业投资基金或个人投资组合等子基金中。随后由子基金对个别初创公司进行投资。这是一种政府资金不直接流向个别企业，而是通过专业基金管理人（GP）流向企业的结构。

《人工智能基本法》将这种现有结构与AI领域相连接，使其可专门用于AI产业。《施行令》第15条之3规定了具体程序。中央行政机关首长可以反映与中小风险企业部长官协商的结果，请求韩国风险投资股份有限公司制定与AI产业相关的投资计划。这意味着科学技术信息通信部长官可以成为提出该请求的主体。

资金来源的构成载于第18条第4款。国家、地方自治团体、公共机构的出资额是基础，在此基础上再加上民间出资者的资金和附带收入。这种设计旨在通过government投入种子资金（seed money），吸引民间资金以数倍规模跟进，从而发挥杠杆效应。

站在初创公司的立场上，这意味着什么呢？如果对AI技术企业进行投资的风险投资基金从母基金的AI专属账户中获得了出资，那么该基金就必须向AI领域进行投资。当AI初创公司尝试吸引投资时，如果寻找此类基金，获得投资的概率就会增加。现有的母基金中也曾设有生物、文化内容、材料零部件等各领域的账户，而《人工智能基本法》则为增设AI账户创造了法律依据。

3. 人工智能研究所：大学与企业共同打造的研究基地

第22条之2为“人工智能研究所的设立及支持等”。这是由2026年1月20日的修改法（第21311号法律）新设，并于2026年1月22日起施行的条款。如果说第12条所规定的人工智能安全研究所是政府主导的安全评估机构，那么第22条之2的人工智能研究所则是民间主导的技术研究基地。两者的性质完全不同。

让我们来看看设立主体。大学、企业等总统令所规定的主体，可以单独或共同设立。《施行令》第16条之2对该范围进行了具体化。这包括《高等教育法》规定的大学、科技领域的政府出资研究机构、非营利法人、《商法》规定的公司、经济与人文社会领域的政府出资研究机构、地方自治团体出资的研究院、特定研究机构，以及科学技术信息通信部长官认定为合适的所有法人或团体。像三星电子这样的大企业、像KAIST（韩国科学技术院）这样的大学，以及民间研究基金会都可以设立人工智能研究所。三星电子和KAIST也可以共同设立。

许可条件有三项：拥有3名以上的发起人；具备执行法定业务所需的人力与设施能力；满足《施行令》规定的追加条件。《施行令》第16条之2第2款作为追加条件，要求具备独立运营所需的财政能力，以及为保护研究数据和防止个人信息泄露的安全对策。在申请许可时，必须向科学技术信息通信部长官提交为期3年的研发项目计划书和资金筹措计划书（《施行令》第16条之3）。

人工智能研究所开展的业务共有6项：AI技术研发、其他技术及跨学科融合研发、研发成果的管理、转让、应用与商业化、专业人才培养、国际交流与国际合作，以及其他研发所需的事项。其中第三项“成果的管理、转让、应用与商业化”十分抢眼。法律将研究所产生的技术转让给企业并转化为产品走向市场的路径，明确规定为研究所的法定业务。

运营资金分为三部分。政府和地方自治团体可以在预算范围内资助经费（第22条之2第6款），可以接受民间的补贴或捐赠（第7款），并可根据章程的规定开展盈利事业（第7款）。其法律形式准用《民法》上的财团法人（第9款）。

在确保人才方面也有相应机制。研究所所长可与大学或企业等协商，接受派遣的员工或允许其兼职工作（第8款）。这意味着KAIST的教授可以兼任人工智能研究所的研究工作，大企业的研究人员也可以以派遣形式来此工作。此外，还可以向派遣员工的所属机构提供必要的支持。

科学技术信息通信部长官拥有监督权限。对于以欺骗或其他不正当手段获得许可的情况，或者无法实现设立目的的情况，必须取消许可。在未满足许可条件、开展目的事业之外的业务、违反法令或章程、侵害公共利益、获得许可后6个月内未启动或1年以上无业绩等情况下，也可以下达纠正命令或取消许可。

同一次修改中并列新设的第22条之3，是科学技术信息通信部长官可以直接设立并运营AI研究机构的依据。其中附带了“为获取创新性人工智能技术而有必要时”这一条件。如果说第22条之2是民间主导，那么第22条之3则是政府主导。该条款使得在民间不愿涉足的领域，如基础研究或与国家安全相关的AI技术等市场

收益不明朗的领域，政府可以直接设立研究所。

与中国的国家人工智能开放创新平台相比，其差异显而易见。中国自2017年起，将百度（自动驾驶）、阿里巴巴（智能城市）、腾讯（医疗影像）、科大讯飞（语音识别）、商汤科技（计算机视觉）等民营大企业指定为国家人工智能平台，并向其集中提供政府课题和数据。这是一种政府选拔并培养“冠军企业”的模式。而韩国的人工智能研究所制度并不限制设立主体。无论是大学、中小企业还是政府出资研究机构，只要具备条件并获得许可，就可以设立研究所。其方向是分散式的。

4. 确保专业人才：国内培养与海外引进

招人是AI初创公司最感到吃力的难题之一。《人工智能基本法》第21条以“确保专业人才”为题，规定了两条人才政策主线。

第一款为国内培养。作为科学技术信息通信部长官可推进的项目，列举了专业人才的岗位能力 education 培训项目、就业支持与促进新人才流入、扩大进入公职的机会、国内外研修与国际交流、改善工作环境与提高待遇。第五项“改善工作环境以及增进待遇等福利提升”在法律中是罕见的表述。这反映了AI人才流失的原因不仅在于薪水，还在于工作条件这一现实。

第三项也引人注目，即“促进专业人才进入公职机会扩大等去向多样化”。这意味着要拓宽道路，使AI专家不仅流向大企业或初创公司，也能进入政府和公共机构。在48个部委厅处已指定副部级首席人工智能官（CAIO）的情况下，需要有实际执行各部门AI相关业务的实务人员。

第二款为引进海外人才。该款列举了从调查和分析海外大学、研究机构、企业的专业人才开始，到构建国际网络、支持海外专业人才在韩就业、国内研究机构走向海外与引进海外研究机构至韩国国内、以及在韩国举办国际组织与国际活动等6项措施。

该条款要真正发挥作用，需要法律之外的政策作为支撑。自2019年起由政府指定的AI研究生院项目便是其中一大支柱。KAIST、高丽大学、成均馆大学、浦项工科大学、蔚山科学技术院等已设立了AI特色研究生院，截至2026年已有10所以上在运营中。在国家人工智能战略委员会第二次全体会议（2026年2月25日）上审议通过的《AI人才培养方向》，是包含99个执行课题的《大韩民国人工智能行动计划》的一大支柱。

让我们与欧盟进行一下对比。欧盟并未在《人工智能法案》（AI Act）中设立专业人才培养条款。相反，它通过“地平线欧洲”（Horizon Europe）框架计划（2021~2027年，总额955亿欧元）和“数字欧洲计划”（Digital Europe Programme，75亿欧元）来支持AI研究人才。AI监管与人才培养运行在不同的法律依据之上。而韩国《人工智能基本法》将监管、振兴和人才培养融入到了是一部法律中。其优点是政策之间的衔接较为容易，缺点是在基本法这一框架下，各项政策的具体针对性可能会被削弱。

5. 公共数据转化为学习资源的路径

训练AI模型需要数据。获取优质数据的成本对初创公司来说是一笔巨大的负担。《人工智能基本法》通过两种路径解决这一问题。

第15条规定了学习用数据相关政策的方方面面。科学技术信息通信部长官必须与关系部委协商，推进旨在促进学习用数据的生产、收集、管理、流通、应用以及确保其质量的政策。政府可以直接实施学习用数据构建项目（第3款），并应当建立综合提供和管理学习用数据的系统（综合提供系统），以供民间自由使用（第4款）。由于规定的是“应当提供”，因此这是义务而非裁量。

第二条路径是2026年修改法新增的第16条第2款第2项之2。其中规定，可以将公共机构拥有和管理的数据生成并提供为学习用数据，并为了确保适当的质量水平而提供必要的支持。这为公共机构将所持有的数据加工并提供给AI学习之用赋予了法律依据。

该条款的实务意义重大。韩国的公共机构拥有庞大的数据。例如，国民健康保险公团的体检数据、国土交通部的交通量数据、气象厅的气象观测数据、特许厅（专利厅）的专利文献数据等。如果将这些数据提供用于学习，医疗AI、交通AI、气候预测AI、专利分析AI的开发成本将大幅降低。虽然仍存在与《个人信息保护法》的冲突、数据质量检验、去标识化处理等实务争议，但法律依据的奠定本身就是一个起点。

《施行令》第12条对学习用数据支持对象项目的评选标准进行了具体化。支持对象包括：学习用数据的生产和加工技术开发、旨在开发AI服务的网络促进及质量确保、指南和标准合同的开发。规定在评选时应同时考虑经济波及效应和是否遵守法令。

6. 为中小企业和弱势群体构建的安全网

《人工智能基本法》第17条虽然简短，却是一个有力的条款。“依据本法实施与人工智能技术及人工智能产业相关的各项支持政策时，应当优先考虑中小企业等。”中小企业、风险企业、小微企业将成为各项支持中优先考虑的对象。在基本计划中也必须反映促进中小企业参与的内容（第2款）。科学技术信息通信部长官可以支持中小企业履行高影响人工智能相关措施并进行影响评估（第3款）。这意味着政府将对遵守监管本身所需的成本提供支持。

第16条第2款第3项规定了对中小企业、风险企业、小微企业员工引进和使用AI技术相关的教育支持。第4项规定了对这些主体引进和使用AI技术所需资金的支持。这为政府直接支持AI引进资金提供了依据。

2026年修改法新增的第17条之2，是向因经济条件难以使用AI产品和服务的人群资助全部或部分费用的条款。《施行令》第1条之2将“AI弱势群体”的范围划分为10种类型。这包括：基础生活保障金领取人与次上位阶层、65岁以上高龄者、残疾人、经历中断女性、求职者、农林渔业从业人员、多文化家庭成员、北韩脱离居民、单亲家庭，以及在使用AI产品和服务方面遇到困难并被中央行政机关或

地方自治团体所认可的人员。对弱势群体的费用资助是原法中没有的制度，是2026年修改法首次构建的安全网。

《施行令》第15条之2进一步扩大了费用资助对象的范围。除了上述类型中的9种以外，还包含了理工科人才、国家有功者中被判定为伤残等级的人员，以及在地方大学就读或毕业的区域均衡人才。这为来自地方大学的青年获得AI服务使用费用的资助开辟了法律途径。

7. 实证基地与集聚区：技术推向市场的路径

来自研究所的AI技术并不会直接投入市场，期间必须经过实证测试。第24条规定，国家和地方自治团体可以构建并运营相关设施和设备，以支持人工智能经营者对其开发或受让的技术进行实证、性能测试和检验认证。公共机构还可以向经营者开放其拥有的实证基础设施。

第23条对人工智能集聚区制度作出了规定。该制度是指通过指定开展AI研发的企业、机构、团体物理性聚集的区域，对其集中提供行政、财政和技术支持的结构。可以通过设立或指定专门机构，对集聚区进行综合支持，并可以出资或补助其运营费用。

第19条的促进融合条款对初创企业也至关重要。该条规定，对于政府作为国家研发项目课题优先反映并开发的人工智能融合产品和服务，应积极予以支持，以确保其能够顺利实施《促进信息通信振兴及融合活跃等相关特别法》规定的临时许可和监管特例实证。这是在AI与医疗、AI与金融、AI与交通相交汇的领域，因现有监管而受阻时，能够通过临时许可寻求突破的法律背景。

8. 扩充财源：资金从何而来

第37条规定了‘为振兴人工智能产业扩充财源等’。国家应制造能够持续、稳定地扩充所需财源的方案，以便有效推进基本计划及本法规定的政策。科学技术信息通信部长官可以建议公共机构对AI振兴项目提供必要的支持。国家和地方自治团体应采取必要措施，以促进民间积极投资AI产业。

该条款具有宣言性，并未写明具体金额或基金规模。实际的财源规模将在每年的预算编制过程中决定。在2026年2月25日国家人工智能战略委员会第二次全体会议上确定的《大韩民国人工智能行动计划（2026~2028）》包含了99项实行课题和326项政策建议，而分配给各课题的预算正是履行第37条扩充财源义务的途径。

9. 初创企业能从本法中获得什么

如果从初创企业的视角来整理整个产业振兴条款，就会清晰地呈现出一幅图景。

在寻找投资资金的阶段，设有创业投资母基金的AI专属账户（第18条第3款）。当技术价值难以证明时，可以期待政府的AI技术价值评估体系（第18条第1款第4项）。在创业初期的教育和培训方面，有创业活跃化教育项目（第18条第1款第2项）提供支持。如果希望使用研发成果，则可以获取政府提供的AI研究及技术开发成果（第18条第1款第5项）。

在数据不足的情况下，提供公共数据作为训练数据（第16条第2款第2项之2）和综合提供系统（第15条第4款）是获取渠道。如果想要培养或获取研究人才，可以使用专业人才培养项目（第21条）和人工智能研究所的人员派遣制度（第22条之2第8款）。AI技术引进资金（第16条第2款第4项）以及优先考虑中小企业（第17条第1款）则有助于减轻初期成本负担。

在产品推出后，实证基础设施（第24条）、监管特例实证（第19条第3款）、优先考虑公共采购（第16条第3款）可以协助其进入市场。在公共采购中优先考虑AI产品和服务，且即便负责该业务的公务员在无故意或重大过失的情况下造成损失也不承担赔偿责任的免责条款（第16条第4款），是旨在促使公务员作出引进AI决定的机制。

If 履行高影响AI义务的成本带来负担，还可以利用科学技术信息通信部长官支持中小企业履行义务和进行影响评估的条款（第17条第3款）。这意味着政府将分担合规成本。

10. 遗留的问题

法律上有依据并不等同于实际上能获得支持。虽然第18条允许利用创业投资母基金，但AI专属账户实际上能分配到多少资金，取决于每年的预算和韩国创业投资的运营计划。作为设立人工智能研究所许可实质性要件的‘自立运营所需的财政能力’具体处于何种水平，仍需等待科学技术信息通信部的公告。提供公共数据作为训练数据如何与《个人信息保护法》、《著作权法》相协调，也是后续需要完善的事项。

在欧盟通过InvestAI提出2000亿欧元投资目标、中国在地方政府层面发放AI补贴的情况下，韩国虽然具备了法律依据，但投入的财源规模目前还无法与之相提并论。

即便如此，有一点是明确的。对于2025年秋天在那家共享办公空间共进午餐的三位代表来说，这部法律并非仅仅给他们增加了监管义务。投资资金的通道、可以获取并使用技术的研究所以及培养和引进人才的制度保障、获取公共数据的途径、进入政府市场的优先权，这些内容都已写入法律。问题在于这些条款能否真正发挥作用。法条已经写就，施行令也已出台。接下来需要预算跟进、许可程序得到完善以及数据真正开放，填补这期间空白的工作目前仍在进行之中。

第七部 违反时会发生什么：制裁与风险管理

第21章. 事实调查的启动条件与资料提交应对

假设某在线平台上散布了人工智能生成的深度伪造（Deepfake）视频。受害者多达数十人，视频通过社交媒体被分享了数万次。科学技术信息通信部要求该平台的人工智能服务提供者提交资料。虽然处于指导期内，但调查已经启动。这就是《人工智能基本法》第40条发挥作用的瞬间。

事实调查是该法的利刃。虽然有批评认为3000万韩元的行政罚款上限与欧盟最高上限3500万欧元相比名存实亡，但给经营者带来实质性打击的并非行政罚款金额。当调查官走进办公室，要求查看账簿和文件的那一刻，经营者的世界便发生了改变。

1. 事实调查的法律结构

《人工智能基本法》第40条在“事实调查等”的标题下，设有三个款项。

第1款规定了启动条件。科学技术信息通信部部长可以启动事实调查的情况有两种。一是发现违规事项或得知存在违规嫌疑的情况。二是接到有关违规的举报或受理了信访的情况。

成为调查对象的违规条款有五项。即生成式人工智能生成物的标识义务（第31条第2款）、难以与真实事物区分的深度伪造等生成物的明确告知或标识义务（第31条第3款）、高性能人工智能的风险识别、评估及缓解义务（第32条第1款）、其履行结果的提交义务（第32条第2款）、高影响人工智能经营者的确保安全性与可靠性措施义务（第34条第1款）。

值得注意的是被遗漏的部分。违反第31条第1款的事先告知义务并非事实调查对象。违反事先告知义务属于第43条直接处以行政罚款的对象。这意味着无需提出提交资料要求或进行现场出入检查，即可处以行政罚款。相反，对于违反生成物标识义务或确保高性能人工智能安全性义务的行为，并不直接处以行政罚款，而是经过事实调查后，若认定违规，则下达中止命令或纠正命令。只有在不履行

该命令时，才会附加行政罚款。

第2款规定了调查手段。所属公务员可以进入经营者的办事处或营业场所，对账簿、文件以及其他资料或物品进行调查。本法中未作规定的事项，遵循《行政调查基本法》。

第3款规定了结果措施。若违规事实得到认定，可以为了中止或纠正相关违规行为而命令采取必要措施。

将该结构压缩成一句话即是如此。认知嫌疑或受理举报是发动条件，提出提交资料要求和现场出入检查是调查手段，中止命令或纠正命令是制裁，而不履行该命令则是处以3000万韩元以下行政罚款的前提。

2. 指导期与例外：何时拔刀出鞘

《人工智能基本法》已于2026年1月22日施行，但在处以行政罚款和事实调查方面设有指导期。这并非法律或施行令中记载的宽限，而是科学技术信息通信部公布的行政执法方针。期限至少为1年以上，且在施行初期优先进行行政指导和纠正劝告。其宗旨是，对于因经营者未认识到义务或不知晓履行方法而产生的违规行为，将选择引导而非惩罚。

指导期并非适用于所有情况。政府表示，在发生人员伤亡事故、人权受损等重大社会问题的极其例外的情况下，即使在指导期内也将进行事实调查。

哪些情况属于此类呢？捕捉到人工智能生成技术被用于制作性剥削物品、名誉毁损、诈骗等犯罪的迹象便是典型情况。人工智能系统的故障或判断错误对人的生命或身体造成直接伤害的情况也是如此。因自动驾驶辅助系统的误判导致事故，或人工智能医疗器械的误诊给患者带来身体伤害的情况均属于此类。人工智能系统在招聘或贷款审查等高影响领域导致系统性歧视的情况亦是如此。

这些例外情况的共同点在于损害的重大性与紧急性。对于遗漏透明度义务或文件化不完善等问题，可以给经营者时间进行纠正，但若是关系到人的生命或犯罪

正在进行中，则没有理由等待。

站在经营者的立场上，这是明确的界限。虽然存在名为指导期的缓冲区，但该缓冲区却破有漏洞。

3. 调查官到来时：事实调查的程序

事实调查一旦启动，科学技术信息通信部部长将使用两种途径。

一是要求提交资料。要求提交相关资料的公文将发送给经营者。二是现场调查。所属公务员将进入经营者的办事处或营业场所。

《人工智能基本法》第40条第2款规定，关于调查的内容、方法、程序，除本法规定的事项外，遵循《行政调查基本法》的规定。这一授权对经营者而言具有双重意义。

《行政调查基本法》具备保护经营者的程序装置。根据《行政调查基本法》第17条，行政机关必须在调查启动7日前以书面形式事先通知调查对象。根据该法第9条至第11条，出席要求书、资料提交要求书、现场出入调查书中必须记载调查的目的、期间、范围、调查员的姓名与职位、需提交的资料、对拒绝调查的制裁及其依据法令。现场调查原则上不得在日出前或日落后进行，但在调查对象同意或在营业场所工作时间进行调查等情况下存在例外。调查员应携带表明权限的证件，并向调查对象出示。

但事先通知存在例外。若判断提前通知可能会导致证据被销毁，可以在调查启动的同时出示调查书，或以口头方式通知调查目的。在深度伪造犯罪等证据可能迅速消失的案件中，有适用该例外的余地。

《行政调查基本法》第4条规定的基本原则同样适用。调查必须在达到目的所需的最少范围内实施。不得为了其他目的而滥用调查权。根据该法第15条，针对同一案件，不得对同一调查对象进行重新调查。仅在获取了怀疑存在违法行为的新证据时除外。通过调查得知的信息，不得用于原本调查目的以外的用途，或提供

给他人。

最后一条款对经营者尤为重要。法律阻断了科学技术信息通信部在调查过程中获取的信息流向其他机构、暴露给竞争对手或向媒体公开的可能性。《行政调查基本法》第4条第5款规定，不得公布行政调查的对象或内容，也不得泄露因职务得知的秘密。

4. 资料提交的范围与界限

在事实调查中，经营者反应敏感的是提交资料要求。“可以要求提交相关资料”的法律措辞较为宽泛。经营者必须拿出的资料范围究竟到哪里为止？

范围由第40条的调查对象条款决定。若嫌疑为违反生成物标识义务，该生成式人工智能服务的标识机制、水印应用记录、提供给使用者的告知画面截图、深度伪造过滤日志将成为提交资料对象。若嫌疑为违反确保高性能人工智能安全性义务，风险识别报告、风险评估结果、缓解措施履行明细、风险管理体系文件即为对象。若嫌疑为违反高影响人工智能经营者责任，风险管理方案、说明方案、使用者保护方案、人为管理监督体系、履行依据文件（5年保管义务）即为对象。

范围看似宽泛，但也有其界限。

《行政调查基本法》第4条第1款设定的比例原则是第一道界限。即达到调查目的所需的最少范围。全面要求提交与嫌疑无关的整体业务资料，便违背了比例原则。在深度伪造标识义务违规调查中，要求提供公司的整体销售资料或人事记录，则超出了比例原则的范围。

资料提交要求书中的记载事项是第二道界限。根据《行政调查基本法》第10条第2款，资料提交要求书中必须记载提交期间、请求提交事由、提交文件、提交文件是否返还、对拒绝提交的制裁以及其他与该行政调查相关的必要事项。对于未记载于要求书中的资料，没有义务进行提交。经营者在收到要求书后，应仔细确认记载事项。若判断提交范围过大，有依据根据调查目的请求将其缩小至必要

的最少范围。

5. 商业秘密与技术秘密的保护

人工智能经营者的核心资产是模型架构、训练数据集的构成、超参数以及提示词工程诀窍。在事实调查中，若该信息被包含在提交对象中，经营者的竞争力可能会在瞬间崩塌。

《人工智能基本法》法律条文中并无正面规定保护商业秘密的条款。但施行令第27条第1款但书规定，高影响人工智能经营者在发布风险管理方案等时，可以排除属于商业秘密的事项。在事实调查局面下，保护的依据来源于《行政调查基本法》和其他法律。

《行政调查基本法》第4条第5款和第6款是第一道防线。不得公布调查内容或泄露因职务得知的秘密，也不得将通过调查得知的信息用于原本调查目的以外的用途，或提供给他入。这既是调查官个人的保密义务，也是机构层面的信息目的限制义务。

《防止不正当竞争及商业秘密保护法》是第二道防线。该法第2条第2项将不为公众所知悉、具有独立经济价值且被作为秘密管理的生产方法、销售方法以及其他对营业活动有用的技术上或经营上的信息，作为商业秘密予以保护。人工智能模型的权重、训练数据策展方法、强化学习奖励函数设计等均可能符合此定义。如果经营者一直将其作为秘密进行管理，且具有经济价值，即可作为商业秘密获得法律保护。

在实务中，经营者可以采取的应对措施有三种。

提交资料的同时明确标示属于商业秘密的信息，并请求保密。如果在要求书所记载的范围中存在属于商业秘密的资料，可以释明该资料为商业秘密，并请求限制阅览范围。本法中并无针对拒绝提交资料本身进行制裁的条款。尽管如此，若不配合调查，将在认定违规的判断中起不利作用，且若根据调查结果下达第40条

第3款的中止命令或纠正命令，不履行该命令将导致被处以第43条第1款第3项的行政罚款，因此，相较于全面拒绝，协商调整范围更为现实。

人工智能基本法支援服务台在此时将发挥作用。由韩国软件产业协会（KOSA）运营的该服务台提供匿名咨询服务。收到资料提交要求的经营者，可就如何在营业秘密与调查配合义务之间划定界限接受咨询。深层法律审查最多需要14天。

6. 欧盟的市场监管机构：更宽的刀，更重的罚

《欧盟人工智能法案》将相当于事实调查的权限赋予了市场监管机构（Market Surveillance Authority）。这与韩国科学技术信息通信部单独行使调查权限的结构不同。

根据《欧盟人工智能法案》第70条，各成员国必须指定一个以上的市场监管机构。该机构负责对高风险人工智能系统进行市场监管。人工智能办公室（AI Office）对通用人工智能（GPAI）模型拥有排他性的监督权限。原本规定中就包含人工智能办公室对同时开发GPAI模型及基于该模型的人工智能系统的经营者进行监督，而于2025年11月提出、2026年7月生效的数字综合修正案（Digital Omnibus）则使该管辖权更加明确。被整合至大型平台（VLOP/VLOSE）的人工智能系统，也因该修正案而受到了人工智能办公室的监督。

市场监管机构的调查权限比韩国更宽。根据第74条，可以访问技术文档、合格评定记录以及训练数据相关记录，在难以通过其他方法进行确认的情况下，甚至可以访问人工智能系统的源代码。根据第79条及后续条款，在必要时可以对人工智能系统进行测试，并可命令在市场上撤回或召回人工智能系统。

制裁的规模有决定性的不同。《欧盟人工智能法案》的罚金因违规类型而异。若违反被禁止的人工智能惯例，最高可处以3500万欧元或全球年销售额7%的罚款，违反其他义务的上限则是1500万欧元或3%。这与韩国的3000万韩元无法相提并论。

韩国看重影响。欧盟看重风险。这种差异在执行阶段也直接贯穿其中。韩国的事实调查重点在于确认违规并命令纠正。欧盟的市场监督则更进一步，旨在将危险的人工智能系统从市场上清除。在韩国，事实调查的终点是纠正命令，若违抗该命令则处以3000万韩元的行政罚款。在欧盟，市场监督的终点是市场撤回命令，一旦违规将面临数百亿韩元的罚金。

中国则以另一种方式加速。中国国家互联网信息办公室（CAC）对生成式人工智能服务赋予事先注册义务（备案），并在注册审查过程中已确保了对技术资料的访问权。各地的互联网信息办公室持续清理未经安全评估的生成式人工智能应用（App）和小程序，2025年6月上海当局删除了82万件非法内容，并使约2700个人工智能智能体（AI Agent）失活。执法的核心并非事后调查，而是事先注册与常态监督。对于违反《个人信息保护法》的行为，最高可处以5000万元人民币（约100亿韩元）或上一年度销售额5%的罚款，但实质上更强有力的制裁是中止服务和停业。

将这三个国家的调查权限排成一行便是如此。韩国是在发现嫌疑后进行调查并命令纠正。欧盟是常态化运行市场监督体系，在违规时将其从市场中清除。中国则是通过事先注册和常态监督防患于未然，在事后则切断服务。

7. 经营者的应对：在调查到来之前

为了应对接受事实调查的可能性，经营者应进行四项检查。

整顿高影响人工智能判断记录。必须保留关于自家人工智能是否属于第2条第4项的十一个领域的事先审查记录。即使判断不属于高影响人工智能，将该依据作为文件保留下来也更为安全。第33条赋予了经营者事先审查是否属于高影响领域的义务。

保管履行依据文件。第34条第1款第5项要求编写并保管能够确认确保安全性与可靠性措施内容的文件。施行令将保管期限规定为5年。在事实调查中，该文件最有可能被优先要求提供。

检查生成物标识体系。由于像深度伪造犯罪这样成为重大社会问题的案件，即使在指导期内也可能启动事实调查，因此，提供生成式人工智能服务的经营者应确认标识义务的履行状态。是否存在肉眼可见的标识、是否留有水印应用日志、向使用者告知系人工智能生成事实的方式是否满足法令的要求等。

构建营业秘密管理体系。针对模型架构、训练数据构成、核心算法，必须采取作为营业秘密进行管理的措施。只有具备了访问控制、秘密标示、保密协议等条件，才能在事实调查中主张对相关资料的保护。未作为秘密进行管理的信息无法获得营业秘密的保护。

需要做好这些准备的理由只有一个。指导期总会结束。此后，对于第40条所列举的义务违反情况便可进行事实调查，违抗纠正命令则将被处以行政罚款。在指导期内建立起体系的经营者，与结束后才匆忙行动的经营者之间的差距，将在第一份事实调查通知书到达的那一天显现出来。

事实调查并非处罚。《行政调查基本法》第4条第4款这样写道：“相较于对违反法令等行为进行处罚，行政调查应将重点放在引导人们遵守法令等方面。”事实调查究竟是揭发与惩罚的工具，还是引导遵守的过程？韩国法律选择了后者。这一选择一贯地融入了3000万韩元这一行政罚款上限、至少1年的指导期以及以不履行纠正命令为前提的行政罚款结构之中。该法向经营者传达的信息并非“惧怕罚款”，而是“建立体系”。

第22章. 行政罚款、纠正命令、罚则体系与辅导期的现实

2026年3月，AI基本法支持中心接到了某初创企业代表的电话。内容是其自家聊天机器人服务在未加入AI事先告知语句的情况下，已运营了两个月。“行政罚款是多少？”咨询员念了法条。首次违反500万韩元。代表沉默片刻后问道：“不过现在是辅导期，所以不会罚钱，对吧？”咨询员的回答很简短。法律上的义务从1月22日起已经生效，辅导期并不意味着法律上的免责。

这一通电话包含了韩国《人工智能基本法》制裁体系的核心。法律已经施行，义务已经生效，且存在行政罚款的规定。然而，政府公开承诺，在至少一年的时间里，将优先进行指导而非处罚。经营者如何解读这一空隙，将决定其应对方式的不同。

1. 三种行政罚款，三种分量

韩国《人工智能基本法》第43条第1款将征收行政罚款的对象限制为三种。法律整体为6章体系，最后一条是第43条，这意味着可处以行政罚款的违规行为只有三种。

第一种情况是未遵守事先告知义务。第31条第1款规定，在提供使用高影响人工智能或生成式人工智能的产品或服务时，必须事先告知使用者“此服务基于人工智能运行”。若遗漏此项，将根据第43条第1款第1项处以行政罚款。从施行令附表2的个别标准来看，首次违反500万韩元，第二次违反1000万韩元，第三次及以上违反1500万韩元。在三种行政罚款中金额最低。事先告知在技术上并不困难。只需在服务界面的某处加入一行文字即可。因此，金额似乎也设定得较低。

第二种情况是未指定韩国境内代理人。根据第36条第1款，在韩国境内无住所或营业所的海外人工智能经营者，若使用者人数或销售额超过施行令标准，必须以书面形式指定韩国境内代理人，并向科学技术信息通信部长官申报。若未指定，从首次到第三次的行政罚款金额均为2000万韩元。未根据次数进行区分。这是

立法者发出的信号，表明将海外经营者规避韩国境内管辖权视为严重问题。因为全球人工智能企业在韩国市场获得数百万使用者，却不设立韩国境内窗口的情况，将使法律执行本身变得不可能。

第三种情况是未履行纠正命令，且金额最重。根据第40条第3款，若未履行科学技术信息通信部长官下达的停止命令或纠正命令，首次违反处以1000万韩元，第二次违反处以2000万韩元，第三次及以上违反处以3000万韩元。在第三次及以上触及法定上限3000万韩元。不履行纠正命令意味着无视行政机关的公共命令，因此受到最严厉制裁。

2. 减免与加重：实际征收的金额会有所不同

施行令附表2的一般标准中包含了减少或增加行政罚款金额的规定。

减免事由有四种。属于《中小企业基本法》规定的中小企业、风险企业或小微经营者的情况；因轻微的疏忽或错误导致违规的情况；被认定为已努力纠正或消除违规状态的情况；其他综合考虑违规行为的种类和程度，认为有必要予以减少的情况。减免幅度在个别标准金额的二分之一以内。例如，违反事先告知义务的首次基本行政罚款为500万韩元，如果是中小企业，最多可减至250万韩元。但是，滞纳行政罚款的经营者不适用减免。

加重事由有三种。违规的内容和程度严重，被认定对使用者造成巨大损害的情况；违规状态持续时间达6个月以上的情况；其他综合考虑违规行为的程度、动机和结果，认为有必要予以增加的情况。加重幅度也在二分之一以内，但不得超过法定上限3000万韩元。若对不履行纠正命令的首次违反（基本1000万韩元）适用加重，最多可升至1500万韩元，而第三次违反（基本3000万韩元）已是上限，因此即使加重也是3000万韩元。

违规次数的计算也有规则。只有在最近2年内因相同的违规行为受到行政罚款处分的情况下，才会适用加重的次数。过了2年，次数就会重置。

3. 事实调查与纠正命令：行政罚款之前的行政手段

在违反生成物标识、确保安全性、高影响人工智能责任的情况中，行政罚款是最后一张牌。在此之前，设有事实调查和纠正命令两个阶段。

第40条第1款规定了科学技术信息通信部长官可以进行事实调查的两种情况。其一，发现违反生成物标识义务（第31条第2款、第3款）、确保安全性义务（第32条第1款、第2款）、高影响人工智能经营者责任（第34条第1款）的事项，或得知相关嫌疑时。其二，接到针对违反上述条款行为的举报或信访时。

事实调查的权限包括要求提交资料和现场调查。根据第40条第2款，所属公务员可以进入经营者的办事处或营业场所，对账簿、文件及其他资料或物品进行调查。调查的内容、方法和程序遵循《行政调查基本法》。

若调查结果认定存在违规事实，可根据第40条第3款，为停止或纠正该违规行为，命令采取必要措施。这就是纠正命令。纠正命令的法律效力属于行政处分。若经营者不履行该命令，才会成为征收行政罚款的对象。

有一点需要注意。违反事先告知义务（第31条第1款）和未指定韩国境内代理人（第36条第1款）不属于事实调查对象，而是直接成为征收行政罚款的对象。生成物标识义务（第31条第2款、第3款）、确保安全性义务（第32条）以及高影响人工智能经营者责任（第34条）则需要经过事实调查，若仍不履行，则走纠正命令、最后行政罚款的三段程序。这两种义务拥有直接进入行政罚款路径的原因，被解释为是否违规在客观上容易确认。

纠正命令可以提起不服。由于是行政处分，可以通过行政复议或行政诉讼进行争议。对征收行政罚款有单独的不服程序。根据《违反秩序行为规制法》，如果在60日内提出异议，将在有管辖权的法院开庭审理。

4. 罚则：唯一可能受刑事处罚的行为

第42条是韩国《人工智能基本法》整体中唯一的刑事处罚条款。

国家人工智能战略委员会的委员若将在职务上知悉的秘密泄露给他人或用于职务以外的目的，处3年以下有期徒刑或3000万韩元以下罚金。依据是第7条第9款的保密义务。

战略委员会是国家人工智能政策的核心审议及表决机构。委员长由总统担任，委员定额60人以内，由18名当然委员和最多41名民间委员组成。他们处理与人工智能相关的国家战略、经营者规制方向、高影响人工智能判断标准等敏感信息。委员若泄露此类信息，不论是否为公务员，都将受到刑事处罚。

第41条是适用罚则时的公务员拟制规定。在适用《刑法》第129条至第132条（贿赂罪）时，委员会委员中非公务员的委员视为公务员。受委托业务的机构或团体的任职员，在此基础上叠加《刑法》第127条（公务上泄密），视为公务员。民间委员若收受贿赂，将承担与公务员同等的刑事责任，若泄露职务上的秘密，不依刑法，而是直接按本法第42条受罚。

法律直接以罚则形式规定的刑事处罚仅有泄密一项，这一事实展现了韩国《人工智能基本法》的性质。对于经营者违反义务的行为仅征收行政罚款（行政罚），而刑事罚仅用于保护治理机构的内部秘密。经营者即便漏办事先告知、拒绝纠正命令或未指定韩国境内代理人，也不会去坐牢。缴纳行政罚款即可了事。

5. 辅导期：法律义务依然存续

辅导期这一表述并未出现在法律条文中。施行令中也没有“辅导期”这个词。这是科学技术信息通信部在2026年1月22日法律施行的同时公布的政策方针。即暂缓征收行政罚款至少一年以上，直到2027年1月之后，在此期间优先进行行政指导和纠正劝告。

该方针的法律依据是科学技术信息通信部长官的裁量权。是否及何时征收行政罚款属于行政机关的裁量事项，因此政府宣布“暂不征收行政罚款”本身并不违法。但是，经营者若将其解读为“因为是辅导期，所以不用遵守义务”，那就危险了。法律上的义务从2026年1月22日起已经生效。未进行事先告知的日子正在

一天天累积。即使辅导期结束后，科学技术信息通信部依然可以根据该期间的违规事实征收行政罚款。因为根据《违反秩序行为规制法》，行政罚款可在违规行为结束之日起5年内征收。

辅导期也有例外。在恶意利用深度伪造犯罪、人员伤亡、大规模侵犯人权等重大事件中，可立即进行事实调查。辅导期只是对轻微程序瑕疵的暂缓，而非对严重损害的免罪符。

在实务上，辅导期向经营者传达了双重信息。既能听到法律已施行、需建立合规体系的信息，又能听到暂时不会处罚、不必着急的信息。在施行10周内，AI基本法支持中心就涌入了552件咨询，这种模糊性是原因之一。在咨询主题中，与透明度相关的占51%，为最多；其次是是否属于高影响人工智能，占19.6%。比起义务本身，经营者对“现在立刻该做什么”感到混乱。

6. 辅导期结束后会有什么变化

辅导期的结束时间尚未正式公告。科学技术信息通信部表示“至少1年”，因此看来不会在2027年1月22日之前结束。不过，这并不意味着1年后会自动结束。预计政府将另行公布结束时间。

辅导期结束后，针对违反事先告知义务的行政罚款征收将成为现实。启动事实调查的可能性也很大。下达纠正命令并追踪履行与否的行政活动预计将全面展开。

从经营者的立场来看，最需要注意的是这种情景：辅导期结束之后，科学技术信息通信部马上开展统一检查。服务界面上没有事先告知语句的经营者、未进行生成物标识的经营者、未具备高影响人工智能风险管理文件的经营者，可能会被集中查处。届时，即使“不知道”或“正在准备中”的抗辩可能被认定为减免事由，但无法抹消违规事实本身。

7. 3000万韩元足够吗

韩国《人工智能基本法》的行政罚款上限为3000万韩元。折合韩元为3000万韩元，约合2万美元。

EU AI Act的最大课征金为3500万欧元或全球年销售额的7%，取两者中较高者。按2026年7月的汇率换算，约为595亿韩元。韩国的3000万韩元与欧盟的3500万欧元之间，有着超过1000倍的差距。

中国对违反各领域AI规定的行为，在适用网络安全法和数据安全法等相关法律法规的罚款的同时，还可并处停业整顿或停止服务命令。2026年1月1日施行的修订版网络安全法大幅提高了罚款上限。有预测指出，在中国正在筹备的综合人工智能法中，罚款上限将更高。韩国的3000万韩元甚至低于中国现行罚款水平。

将三个国家并列来看，制裁强度的顺序非常清晰。欧盟的3500万欧元排在最上面，除了罚款还要加上停业整顿的中国排在中间，韩国的3000万韩元排在最下面。如果中国通过综合法，差距可能会进一步拉大。

从法律制定之初，就有人提出“3000万韩元这个金额能否给大企业带来遵守动机”的问题。在第22届国会科学技术信息广播通信委员会审议过程中，当行政罚款水平成为争议焦点时，“韩国AI产业正处于初期成长阶段，过度的罚则会萎缩创新”的逻辑为低金额提供了正当性理由。对于年销售额数万亿韩元的企业来说，3000万韩元类似于违章停车罚单。如果通过违规行为获得的利益大于行政罚款，那么缴纳行政罚款并继续违规将成为合理的选择。

也有反对意见。韩国并非只用行政罚款进行规制。在行政罚款之前，还设有事实调查、纠正命令、行政指导和指南劝告。拒绝科学技术信息通信部的劝告，甚至无视纠正命令的经营者，将承担业界声誉受损、被排除在公共采购之外以及因后续立法加重制裁等间接成本。其逻辑是，3000万韩元这个数字并非全部。

然而，间接成本不具法律强制力。欧盟凭借销售额7%这一金额本身就足以改变企业的行为。中国则通过停业整顿这一对事业存续的威胁来促使企业顺从。韩国

目前还没有那样的工具。

8. 欧盟：根据违规类型分级的三个阶段课征金

EU AI Act的课征金体系根据违规的严重程度分为三个阶段。

违反被禁止的人工智能惯例（第5条）的情况最为严重。最高金额为3500万欧元或全球年销售额的7%，取两者中较高者。社会信用评分、潜意识操纵人工智能、无差别的生物识别监控等行为均属此列。韩国没有这种源头禁止条款，因此不存在会被征收这种水平课征金的违规行为。

违反高风险人工智能义务和GPAI义务的情况，最高为1500万欧元或全球年销售额的3%。

向当局提供不准确信息的情况，最高为750万欧元或全球年销售额的1%。对于中小企业，适用各自金额中较低的一项。

与韩国3000万韩元的单一上限相比，欧盟等同于根据违规类型对金额进行了分级分配。体现了比例原则：对高风险违规行为处以重罚，对程序性违规行为处以轻罚。

欧盟通过2025年11月执行委员会提议并于2026年上半年确定的数字总括（Omnibus）修订案，将独立高风险人工智能义务的适用时间推迟到了2027年12月2日。内嵌于产品的高风险人工智能将从2028年8月2日起适用。虽然期限推迟，但制裁水平并未改变。其功能类似于韩国的辅导期，但差异在于，欧盟是通过修改法律来变更施行日期本身，而韩国则是在保留法律的同时通过行政裁量暂缓执行。

9. 中国：比罚款更可怕的停业整顿

中国的人工智能制裁不仅限于金钱罚款。

根据《生成式人工智能服务管理暂行办法》（2023年8月施行），如果服务提供者未履行事先安全评估或算法备案，将受到警告、通报批评、限期责令纠正、暂停提供相关服务的措施。如果同时违反《数据安全法》或《网络安全法》，除了根据相应法律处以罚款外，还可能下达停业整顿、吊销营业执照、停止服务的命令，且2026年1月1日施行的修订版网络安全法大幅提高了罚款上限。国家互联网信息办公室（CAC）根据2025年“清朗”整治计划，相继下架了未经安全评估的生成式人工智能App和小程序。在2025年6月上海的“亮剑浦江2025”专项行动中，下架了54个不合格的生成式人工智能App，并有3起被刑事立案。

韩国的《人工智能基本法》中没有停业整顿或停止服务命令。虽然可以下达纠正命令，但由于使用的是“为停止或纠正该违规行为所需采取的措施”（第40条第3款）的表述，这与中止业务本身有所不同。

中国正在准备的综合人工智能法尚未公开正式草案。有预测称，方向将是提高罚款上限，并强制要求记录训练数据来源和对大模型进行安全影响评估。

比较三个国家的制裁手段，可以看出结构不同。欧盟以巨额课征金切断违规的经济诱因。中国在罚款之上，还增加了停业整顿和停止服务这种对事业存续的威胁。韩国只有行政罚款和纠正命令，且金额上限较低。

10. 规制重新研讨与后续立法的可能性

施行令第33条赋予科学技术信息通信部长官重新研讨规制的义务。必须以2026年1月1日为基准，每三年审查高性能人工智能安全性确保义务对象标准、指定韩国境内代理人经营者标准等的妥当性，并采取改进措施。但是，行政罚款并不在这个重新研讨的对象之内。征收标准需要通过修改施行令附表的方式，而3000万韩元的上限只能通过修订法律第43条的方式来改变。

国会立法调查处研究官郑俊和在2025年3月的《焦点与论点》（第2327号）中指出，需要根据人工智能经营者的类型区分义务。据悉，韩国法制研究院研究委员郑元俊也在2025年9月的报告中发表了制裁水平不足以确保规制实效性的见解。

学术界和政策研究机构双方均有意见认为，3000万韩元的上限是法律制定当时政治妥协的产物，随着产业的成长，上调是不可避免的。

欧盟和中国的立法动向也会对韩国产生影响。欧盟虽然通过总括（Omnibus）修订案将高风险人工智能义务的施行推迟到2027年12月，但维持了课征金水平。如果中国通过综合人工智能法，韩国很有可能成为唯一一个制裁水平显著偏低的国家。国际规制协调（harmonization）的压力可能会加速韩国上调行政罚款。

11. 经营者现在该做什么

辅导期是暂缓而非免除。法律上的义务已经生效，违规事实会留下记录。

必须从事先告知开始。如果利用高影响人工智能或生成式人工智能提供服务，必须在使用者使用服务之前，告知其该服务基于人工智能运行的事实，以便其知晓。仅仅在屏幕上加入这句话，就可以避免首次违反500万韩元的行政罚款。

必须检查生成物标识。生成式人工智能的产出物必须带有AI生成标识，对于难以与真实事物区分的深度伪造级别的产出物，必须以使用者能清晰识别的方式进行告知或标识。仅有不可见水印是不够的。

如果是海外经营者，必须确认是否指定韩国境内代理人。如果不指定，不分次数，行政罚款都是2000万韩元。

如果是运营高影响人工智能的经营者，必须具备风险管理文件、说明方案、使用者保护方案和管理监督体系，并将相关文件保存5年。辅导期结束后，一旦开始事实调查，这些文件的存废将成为决定是否下达纠正命令的核心证据。

3000万韩元对大企业来说可能是个小数字。然而，行政罚款征收记录、纠正命令记录和行政调查记录将影响参与公共采购、进军海外市场以及招商引资。记录比数字更沉重。

辅导期终会结束。在结束前准备妥当的经营者和结束后才行动的经营者之间，会产生一次行政罚款的差距。那一次所留下的行政记录，存在的时间将比想象的更长。

第23章. 应对监督的文件管理——应当保管哪些文件，保管多少？

2026年4月，首尔江南区的一家金融科技公司接到了一起关于贷款审查人工智能的投诉。内容是一名信用评分为750分的30多岁上班族被拒绝贷款，而同一信用评分的另一名申请人却获得了批准。科学技术信息通信部根据韩国《人工智能基本法》第40条启动了事实调查。调查官向该公司要求的第一份资料是记录人工智能模型判断依据的文件。该公司拿出了风险管理方案文件。那是为了应对2026年1月法律施行而匆忙制作的2页文件。其中既没有风险识别明细，也没有训练数据概要，更没有异议处理记录。调查官问道：「根据《施行令》第27条第2款的规定，履行依据文件在哪里？」该公司的回答是「还在整理中」。

1. 法律所要求文件的核心框架

韩国《人工智能基本法》第34条第1款要求高影响人工智能经营者履行六项措施。其中第5项就是文件。条文原文如下：「编写并保管能够确认确保安全性和可靠性措施内容的文件。」法律所称的「文件」是证明安全性和可靠性措施整体的记录。并非仅记录风险管理。这意味着要制作并保管涵盖第34条第1款其余五项，即风险管理方案（第1项）、说明方案（第2项）、使用者保护方案（第3项）、人员管理监督体系（第4项）以及委员会审议表决事项（第6项）在内的全部内容的文件。

保管期限在《施行令》第27条第2款中予以了明确规定，为5年。括号内写明包含通过电子方式进行的保管。这并非要求打印成纸质文件放入文件柜。电子文件、数据库、云存储均被允许。

5年的起算点在《施行令》中并未另行明确。从措施履行之时起算是一种稳妥的解释。如果风险管理方案是在2026年3月制定的，那么可以认为应保管至2031年3月。如果在此期间对方案进行了修订，则修订前的版本和修订后的版本应分别保管5年，这是稳妥的做法。在这一结构下，版本管理实际上成为必然要求。

2. 六种保管对象文件

下面逐一梳理与第34条第1款各项相对应的文件。

风险管理方案文件（第1项）涵盖风险管理政策、组织体系，以及风险识别、分析、评估、处理和改进的全过程。《施行令》第27条第1款第1项要求在官方网站上公示「风险管理政策及组织体系等风险管理方案的主要内容」。除了用于公示的摘要版本外，内部保管用文件必须包含已识别的风险列表、各风险的评估分数、缓解措施的履行日程以及对残余风险的判断依据。

说明方案文件（第2项）在技术上可行的范围内，记录人工智能得出的最终结果、得出结果所使用的主要标准，以及用于开发和训练的训练数据的概要。由于附加了「在技术上可行的范围内」这一但书，这并不意味着要针对黑盒模型将所有推理路径逐一详尽写出。然而，记录应用SHAP、LIME等可解释性工具的结果、主要输入变量的权重信息，或者训练数据的来源、规模及偏差检查结果等内容，是稳妥的做法。

使用者保护方案文件（第3项）涵盖异议提出程序、错误纠正程序以及损失救济方法。如果是通过高影响人工智能审查贷款或筛选招聘候选人的服务，必须写明被拒绝的使用者应在何处以何种方式提出异议，以及公司将在几个工作日内通过何种程序进行审查。

人员管理监督记录（第4项）包含负责管理和监督该高影响人工智能的负责人的姓名和联系方式（《施行令》第27条第1款第4项）、监督活动日志以及人工干预实际发挥作用的案例。如果有人工智能做出异常判断时人工介入并更改结果的记录，这正是第4项的履行证据。

关于文件编写与保管本身的记录（第5项），是管理上述四种文件的编写日期、修订历史、保管位置及访问权限设置的元数据文件。需要有一份文件来证明保管文件这一事实本身。

委员会审议表决事项（第6项）适用于国家人工智能战略委员会为确保高影响人工智能安全性和可靠性而另行进行审议表决之内容的情形。截至2026年7月，尚未公开基于该条款的具体表决事项，但今后如果委员会决定了特定行业的安全标准或针对特定技术的附加要求，相应的履行记录也将成为保管对象。

3. 公示与保管不同

必须区分《施行令》第27条第1款与第2款的差异。第1款是「公示」。即在事务所、营业场所以及互联网官方网站上发布并向外部公开。第2款是「保管」。即在内部留存记录。

公示对象仅限于四项。即风险管理方案的主要内容、说明方案的主要内容、使用者保护方案以及管理监督负责人的姓名和联系方式。因为规定的是「主要内容」，所以没有义务上传全文。

属于《防止不正当竞争及商业秘密保护相关法律》第2条第2项规定的商业秘密的事项，可以从公示中排除。对于模型架构的详细结构、独有的训练数据处理技术、权重参数等信息，可以主张商业秘密。但在保管义务上没有此类例外。即使是商业秘密，也必须在内部制作文件并保管5年。

保管对象为第34条第1款的全部内容，即六项措施的全部履行依据文件。其结构是选择性地公开进行公示，而毫无遗漏地记录进行保管。

4. 电子文件的格式与要求

既然《施行令》明确规定「（包含通过电子方式进行的保管）」，那么在实务中就没有理由固守纸质文件。问题在于，法律并未对电子文件的格式设定具体标准。

科学技术信息通信部的《高影响人工智能经营者责任指南》并不强制规定文件格式。无论是PDF、Word文件、电子表格、内部Wiki，还是文件管理系统（DMS），任何形式只要处于「能够确认措施内容」的状态即可。

实务中推荐的格式原则有三个。

可检索性。在5年后进行事实调查时，必须能够立即找出该时点的文件。如果仅以扫描图像进行保管，检索会很困难。经过OCR处理的PDF或结构化的数据库会更好。

完整性证明。必须能够证明文件在事后未被篡改。自动记录编写日期和时间的系统、留有修改历史的版本管理工具、哈希值记录均属于此列。也有企业甚至引入基于区块链的时间戳，但这并非法律上的强制要求。

访问控制。对于属于商业秘密的文件，必须限制访问权限，方能维持商业秘密保护要件（不为公众所知悉、经济实用性、保密管理性）。如果不保留访问日志，日后主张商业秘密可能会变得困难。

5. 保管系统构建实务

虽然知道必须制作文件，但应该存放在哪里、如何存放，法律却保持沉默。在实务中，可以采取三种级别的处理方式。

最低级别是在现有的内部文件服务器或云存储（如Google Drive、Microsoft SharePoint等）中建立文件夹结构进行保管的方式。这种方式几乎不产生费用。只要制定文件命名规则（例如：「风险管理方案_v2.1_20260315」），在每次修改时保存为新版本并设置访问权限，就能满足法律上的最低要求。问题在于，5年过去后文件夹可能会散落各处，如果负责人离职，就会变得谁也不知道什么文件在哪里。

中等级别是使用全公司范围的文件管理系统（DMS）或电子审批系统的方式。例如Hancom的Hancom Docs、Cyberdigm的Cloudium、微软的Microsoft Purview等产品。在注册文件时自动进行版本编号、记录修改历史，如果设置了保存期限（retention policy），就能防止在5年到期前被误删的事故发

最高级别是引入GRC（Governance, Risk, Compliance，即治理、风险、合规）集成平台的方式。ServiceNow GRC、Archer IRM、OneTrust等平台将风险登记册、控制项、审计轨迹（audit trail）以及报告生成纳入统一体系中进行管理。如果还能自动收集人工智能模型的部署历史、监控日志以及人工干预记录，文件编写的负担就会减轻。但相应的，引入成本似乎高达数千万至数亿韩元。

无论选择哪种级别，核心在于5年后能否在调查官面前拿出这些文件。系统是否华丽并不重要。只要记录完好无损地保留着，能够证明是何时制作的，如果是中途修改的能够追踪修改了什么，这就足够了。

6. 人工智能日志与文件的边界

人工智能系统在运行过程中会生成大量日志。推理请求日志、输入数据、输出结果、响应时间以及错误代码每一秒都在累积。这就产生了一个问题：这些日志是否就是第34条所指的「文件」？

法律要求的文件是「能够确认措施内容」的记录。必须能够确认制定了风险管理方案的事实、实施了说明方案的事实以及人员进行了监督的事实。原始日志（raw log）是系统运行的痕迹，其本身并不是「措施的内容」。

然而，在某些情况下，如果没有原始日志，就很难证明这些措施。「人员管理监督」（第4项）要得到证明，就需要有在特定日期人工审查并修改人工智能判断的记录。这一记录提取自运行日志。要主张「履行了风险缓解措施」，监控系统检测到异常迹象并发出警报的记录就成为了依据。

总结如下：没有义务将全部原始日志保管5年。但是，将证明履行措施所需的日志提取出来包含在文件中，或者单独进行保管，是稳妥的做法。建议将安全性检查日志、人工干预日志、异常迹象应对日志列入保管对象。虽然没有必要将日常的推理请求日志也全部积累5年，但在贷款审查或招聘等高影响领域，从应对事实调查的角度来看，保管个别判断的依据记录是明智之举。

7. 事实调查一旦启动会发生什么

第40条赋予科学技术信息通信部部长两项权限。即要求提交资料和现场调查。

要求提交资料是指要求经营者提供相关资料的书面请求。当发现高影响人工智能经营者违反责任（第34条第1款）或有违反嫌疑，或是接到举报或投诉时，将启动该程序。违反文件化义务也包含在事实调查事由中。如果没有文件，可能会被推定为未履行措施。

现场调查是指所属公务员进入事务所或营业场所，对账簿、文件、其他资料或物品进行调查。程序遵循《行政调查基本法》。原则上须在调查前7日内发出书面通知，必须出示调查员证，并可提出保护商业秘密的要求。

调查结果若认定存在违规行为，科学技术信息通信部部长可以命令采取中止违规行为或纠正违规行为所必需的措施（第40条第3款）。如果不履行中止命令或纠正命令，将被处以3,000万韩元以下的罚款（第43条第1款第3项）。

在实务上需要防范的事项很明确。应将调查官要求提交文件时在30日内提交作为内部目标，并预先整理好文件位置和访问方法。即使负责人离职，继任者也必须能够找到文件。如果文件散落在多个部门，仅是将它们集中到一处可能就需要数周时间。指定人工智能合规负责人，并由该负责人管理所有文件的列表和位置，是基础的防范措施。

8. 欧盟AI法案的技术文件：10年以及合格评定

欧盟AI法案针对高风险人工智能系统的技术文件（technical documentation）规定了更为详尽和沉重的要求。

第11条（Article 11）要求高风险人工智能系统在投放市场或投入服务之前编写技术文件。该文件的目的在于证明该系统符合法律规定的要求，并且必须包含足够的信息以便国家监督机构（national competent authorities）和认证机构（notified bodies）进行评估。

技术文件的内容在附件四（Annex IV）中有明确规定。包括系统的一般说明（开发意图、开发者信息、版本）、系统要素的详细说明（架构、算法、数据处理程序）、有关训练数据的信息（数据收集方法、预处理、标签、统计特征）、关于监控、功能与控制的说明、风险管理系统的实施信息、关于合格评定程序的记录，以及整个生命周期中的变更事项。

保管期限为10年。第18条第1款（Article 18(1)）规定，高风险人工智能系统的提供者（provider）须自该系统投放市场或投入服务之日起将技术文件保管10年。通用人工智能（GPAI）模型也必须根据第53条（Article 53）编写技术文件并保持更新。不过，GPAI技术文件保管10年并非条文规定的义务，而是2025年7月公布的GPAI行为守则（Code of Practice）所设定的自愿遵守标准。

韩国是5年，欧盟是10年。单看数字相差一倍。但真正的差异不在于期限，而在于文件的用途。欧盟的技术文件是合格评定（conformity assessment）的依据资料。要想将高风险人工智能推向市场，必须事先通过合格评定，并且监督机构可以随时要求提供技术文件进行重新验证。韩国的文件是事后事实调查的证据资料。没有事先提交的义务，由监督机构在存在违规嫌疑时提出要求。

对于既在欧盟开展业务又在韩国开展业务的经营者来说，其实务启示很明确。按欧盟标准制作技术文件，就能自动满足韩国要求。反之，如果只满足韩国要求，在进军欧盟时就得从头重新制作文件。如果瞄准这两个市场，从一开始就按欧盟标准制作文件，并将保管期限设定为10年，这样更为高效。

9. 中国的备案与记录义务

中国以算法备案制度为核心赋予文件义务。根据《生成式人工智能服务管理暂行办法》第17条规定，具有舆论属性或者社会动员能力的生成式人工智能服务，在向公众提供之前，必须进行安全评估，并向国家互联网信息办公室（CAC）进行算法备案。备案后，如果算法的原理或功能发生重大变更，也必须在10个工作日内更新备案信息。

2025年9月施行的《人工智能生成合成内容标识办法》要求保留双重标识（显式标识和元数据或数字水印）的履行记录。强制性国家标准（GB 45438-2025）规定了标识方法的技术规范。

与韩国的差异在于是否需要事先备案。在中国，如果属于备案对象范围内的服务，要在开始之前向当局提交文件。而在韩国，是在开始服务后于内部保管文件，待当局要求时再拿出来。编写文件的负担差不多，但提交的时间点不同。

将这三个国家/地区放在一起看，文件义务的强度非常明显。欧盟最重。有事前合格评定依据、保管10年、附件四的详细清单。中国其次。有事前备案、重大变更时更新、遵守国家标准。韩国最轻。事后保管、5年、格式自由。

10. 双重义务与重复监管的放宽

许多高影响人工智能经营者已经承担了其他法律规定的文件义务。金融公司根据《电子金融交易法》和《信用信息的利用及保护相关法律》，将计算机数据保存5年以上。医疗器械制造商根据《医疗器械法》保管设计记录和质量记录。处理个人信息的经营者根据《个人信息保护法》保管处理记录。

韩国《人工智能基本法》第34条第3款和《施行令》第27条第5款（附件1）放宽了这种重复。如果已根据其他法令履行了相当于第34条第1款各项的措施，则视为履行了《人工智能基本法》上的措施。

例如，如果运营贷款审查人工智能的金融公司遵守了《信用信息法》第35条之2规定的说明义务，并具备了该法第36条之2第1款规定的说明义务履行程序，则被视为履行了韩国《人工智能基本法》第34条第1款第2项（说明方案）。使用者保护方案（第3项）在完全履行《金融消费者保护相关法律》第10条规定的责任时可以获得认可。相反，风险管理方案（第1项）的认可对象中不包括贷款审查，因此必须另行履行。具体哪些法律的哪些措施能够获得认可，已在《施行令》附件1中按法律进行了整理。

然而，放宽重复并不等于豁免。如果依据其他法律采取的措施无法涵盖第34条的所有项，那么其余部分必须另行履行并形成文件留存。即使金融公司根据《信用信息法》使其说明方案获得了认可，第1项的风险管理方案、第4项的「人员管理监督」以及第5项「关于文件编写与保管」本身的记录，依然作为韩国《人工智能基本法》的固有要求保留下来。

11. 文件管理自动化的现实

市场上已经出现了自动生成和管理人工智能合规文件的工具。

模型卡（Model Card）和数据表（Datasheet）是Google和Microsoft研究团队提出的标准化文件格式。它们在一页中梳理了模型的目的、训练数据、性能指标、偏差评估结果以及预期使用范围和不适当的使用范围。像Hugging Face这样的模型中心，就提供了自动生成模型卡的功能。

MLOps流水线工具（如MLflow、Weights & Biases、Neptune等）可自动记录模型训练的超参数、数据集版本、性能指标和部署历史。这一记录可以作为第34条第2项（说明方案）的依据文件使用。由于每次运行训练时都会自动留下实验记录，因此事后无需去重建「使用什么数据、是如何训练的」的过程。

OneTrust、BigID等隐私和人工智能治理平台，会集中管理人工智能资产清单（组织运行的人工智能系统列表）、风险评估以及影响评估结果。由于其中许多功能是为应对欧盟AI法案而设立的，若要符合韩国法律的要求，需要进行一些定制调整。

也有开源工具。Google的SAIF（Secure AI Framework）清单和NIST AI RMF（Risk Management Framework）的自我评估工具就是代表。它们可以在没有成本负担的情况下启动，但引入和维护需要人力。

无论是引入工具还是用Excel进行管理，决定文件是否满足法律要求的不是工具，而是内容。自动化工具不会为你填补空白。识别和评估风险是人的判断，而工

具只是记录这种判断的手段。

12. 中小企业的现实应对

要求只有20名员工的人工智能初创公司引入GRC平台是不现实的。我们来探讨一下中小企业如何以最低的成本和人力满足法律要求的方法。

首先必须确认自家人工智能是否属于高影响类别。如果不属于第2条第4项规定的11个领域（能源、医疗保健、招聘、贷款审查、交通、公共服务、学生评估等），那么第34条规定的文件义务本身就不适用。如果对是否属于高影响不确定，可以向科学技术信息通信部提出确认请求（第33条第1款）。确认请求是免费的，并在30日内得到答复。

如果属于高影响类别，则必须制作最低限度的文件集。在一个Google电子表格中创建风险登记册，按风险项目记录识别日期、严重程度、发生可能性、缓解措施以及缓解状态。在一份Word文件中写下说明方案和使用保护方案。梳理训练数据的来源、规模和预处理方法。指定管理监督负责人，并在官方网站上公示姓名和联系方式。将这些文件按日期存放在云存储的文件夹中，每次修改时保存为新文件，就能实现版本管理。

AI基本法支持服务台（swai.or.kr/AI_act_helpdesk）会免费解答这类实务问题。在施行短短10周内就接到了552件咨询，并且还可以进行匿名咨询。深入的法律审查最多需要14天。《高影响人工智能经营者责任指南》也已免费公开。

如果是从开发经营者处接收并使用人工智能的使用经营者，应当查阅《施行令》第27条第3款。如果开发经营者已经履行了第34条第1款第1项至第3项的措施，并且使用经营者没有显著改变该人工智能的原有目的或用途，则认定使用经营者在该范围内履行了措施。即便在这种情况下，保管能够确认履行事实本身的文件（例如：开发经营者的风险管理履行确认书副本）也是一种稳妥的做法。《施行令》第27条第4款规定，使用经营者可以要求开发经营者提供履行第34条所需的资料。

13. 没有文件会怎样

韩国《人工智能基本法》关于行政罚款的规定（第43条）并没有直接将违反文件保管列为处以行政罚款的事由。行政罚款事由仅限于三项。即违反透明度义务（第31条第1款）、未指定韩国境内代理人（第36条第1款）、不履行中止命令或纠正命令（第40条第3款）。

即便如此，也不意味着可以无视文件保管。其路径是这样的。如果在根据第40条进行的事实调查中发现文件不完备，可能会以违反第34条第1款为由下达纠正命令。如果不履行纠正命令，就会被处以行政罚款。因为没有文件这一事实直接成为未履行措施的证据，所以文件保管实际上是履行所有义务的前提条件。

政府表示在法律施行后将设立至少1年的指导期。根据这一方针，对于遗漏透明度义务或文件化不完备的情况，将优先进行行政指导和纠正建议。恶意用于深度伪造犯罪、造成人员伤亡事故以及发生大规模侵犯人权等重大事件的情况则是例外。

指导期结束后情况就不同了。如果没有文件，就无法证明是否进行了风险管理、是否制定了说明方案、人员是否进行了监督。在事实调查中，「履行了但未记录」的主张很难被接受。没有记录就没有支撑履行的证据，调查结果难免会向不利于经营者的方向倾斜。

14. 韩国境内代理人的文件检查义务

如果海外人工智能经营者在韩国开展业务并指定了韩国境内代理人，韩国境内代理人也有着与文件相关的职责。第36条第1款第3项在将「为履行第34条第1款各项规定的确保安全性和可靠性措施所需的支援」纳入韩国境内代理人的业务范围时，在括号内明确标注「包括对同款第5项所规定文件的更新性和准确性的检查」。

韩国境内代理人必须检查海外总部制作的文件是否为最新且准确的。这并非形式上的确认，而是实质性的审查。必须确认风险管理方案是否反映了韩国市场的

状况，说明方案是否适合韩语使用者，管理监督负责人信息是否目前仍然有效。

从海外经营者的立场来看，因为必须与韩国境内代理人共享文件，所以必须先签订保密协议（NDA）。在与韩国境内代理人共享属于商业秘密的技术信息时，必须明确保密义务。

15. 文件就是证据

我们再回到2026年4月首尔江南区的那家金融科技公司。当调查官要求提供风险管理方案文件和履行依据记录时，如果公司拿出了系统的文件会怎样呢？如果在风险登记册中已识别出贷款拒绝偏差风险，并且在缓解措施中记录了特定变量的权重调整及定期的偏差审计，同时还附有最近的审计结果，调查官就可以判断「这家公司正在履行第34条规定」。也可以通过文件来解释，投诉的原因并非系统性偏差，而是个案的差异（收入结构、负债比例等）。

文件管理不仅是防范监督的防御手段，同时也是经营者自行检查自身系统的工具。在将风险整理成文件的过程中，能够发现一些漏洞。在编写说明方案时，也可能会直面「我们也不知道这个模型为什么会得出这样的结果」这一事实。而这种直面正是改进的开始。

5年的保管期限意味着在此期间即使技术变了、负责人变了、商业模式变了，记录也必须留存。在人工智能越来越多地代替进行判断的时代，记录并保管这种判断的依据已成为法律义务。法律所追问的归根结底只有一个问题：你的人工智能为什么会做出那样的判断？你能否在5年后依然能够拿出对这个问题的答案？

第八部 海外扩张与外国竞争者的读法

第24章. 与EU AI Act并列看待的韩国法律

2025年2月2日，欧盟全境禁止了社会评分人工智能系统。该技术是指政府根据市民的在线行为和社会声誉进行评分，并根据该分数限制其获取公共服务。这是EU AI Act第5条生效的日子。与此同时，首尔一家人工智能初创企业代表向法务团队询问，其招聘面试分析服务若要进入欧洲市场需要做哪些准备。等了3周才得到回复。原本以为只需遵守韩国法律即可，但回复内容是，在欧盟法律面前可能需要重新设计整个商业模式。

韩国《人工智能基本法》与EU AI Act在同一时期规制相同对象，但从立法的哲学起就有所不同。其差异横跨词汇选择、分类体系、事前控制方式、制裁水平、禁止范畴、治理结构等六个层面。

1. 风险（risk）与影响（impact）：两部法律的分类哲学

欧盟看重风险。韩国看重影响。

EU AI Act将人工智能系统分为四个层级。不可接受风险（Unacceptable Risk）被从源头禁止。高风险（High-Risk）必须经过事前符合性评估才能进入市场。有限风险（Limited Risk）附带透明度义务。最小风险（Minimal Risk）则可自由运营，无额外规制。从金字塔顶端到底部，这是一种义务随风险强度而变化的结构。

韩国《人工智能基本法》设定了“高影响人工智能”这一单一范畴。其对象是可能对人的生命、身体安全及基本权利产生重大影响或引发风险的人工智能系统（第2条第4项）。能源供应、医疗保健、招聘、贷款审查、交通、公共服务、学生评价等十个领域被直接列举在法律第2条第4项中，并附加了一个可通过施行令增加领域的授权条款。对于非高影响的人工智能，除了透明度义务以及附加于满足大规模运算量标准系统的确保安全性义务（第32条）外，不受其他规制。

词汇的选择背后存在政治考量。在科学技术信息广播通信委员会审议过程中，初期法案中也曾出现过“高风险”的表述。李海旻议员对此表示反对。他发言称：“全球之所以走向‘影响’，是因为影响不仅能包含风险，也能包含范围。”这意味着不愿对技术本身贴上污名标签。当欧盟用“high-risk”一词聚焦于潜在危害时，韩国则用“high-impact”一词将视线转移到了社会影响力上。

这一选择也伴随着代价。韩国法制研究院研究员郑元准在2025年9月的报告中指出，“影响”和“风险”在法条中并存规制，从而引发了混乱。高影响人工智能的定义条款本身就混入了“可能引发风险”的表述。虽然宣称要看重影响，但实际在判断标准中又重新引入了风险。欧盟细分为4个层级并分配了适合各个层级的义务，而韩国在单一范畴内对所有高影响人工智能赋予相同的义务。这正是韩国法制研究院和学术论文（白水源，《欧洲宪法研究》第48期）中反复提及这种单一结构局限性的原因。

2. 事前符合性评估 vs 自主影响评估

在EU AI Act中，若要将高风险人工智能系统投放市场，必须通过符合性评估（conformity assessment）。属于产品安全规定（附件I）的情况，需要第三方认证机构（notified body）的评估。属于附件III的独立型高风险人工智能可进行自我符合性评估，但必须具备技术文件、质量管理体系以及欧盟符合性声明。未通过符合性评估将无法获得CE标志，没有CE标志则不能在欧盟市场销售。

韩国法律中没有这种事前关卡。赋予高影响人工智能经营者的义务包含在第34条第1款这一条文的五个项中：树立与运营风险管理方案（第1项）、树立与施行说明方案（第2项）、树立与运营使用者保护方案（第3项）、人的管理与监督（第4项）、文件的编制与保存（第5项）。保存期限5年由施行令第27条第2款规定。在此基础上加上基本权利影响评估（第35条第1款），但这并非强制义务，而是“自主努力义务”。科学技术信息通信部指南虽然提出了影响评估的3个阶段（事前准备、正式评估、事后措施），但该指南本身不具有法律约束力。

所谓自主并非意味着毫无强制力。履行影响评估的经营者在政府采购中将获得优先考虑。法律第16条第3款规定，国家机关等在购买业务所需产品或服务，或者发包服务时，应优先考虑科学技术信息通信部部长确认的人工智能产品和服务，该确认的标准和程序由公告规定（施行令第15条第4款、第5款）。第35条第2款规定，国家机关等欲使用利用高影响人工智能的产品或服务时，应优先考虑实施了影响评估的产品和服务。虽然法律本身不强制进行影响评估，但政府采购这一市场准入条件实质上引导了评估，属于一种“软性规制”。

将两种体系并列来看，方向十分清晰。欧盟说：“在进入市场之前证明安全。”韩国说：“进入市场之后自行管理，若想进入政府市场请展示评估结果。”欧盟的方式有利于消费者保护，但创新速度可能会减慢。韩国的方式减轻了经营者的初期负担，但也可能变成问题爆发后才应对的“马后炮”。这两种方式都不完美。问题在于要承受何种失败。

3. Provider/Deployer vs 开发经营者/使用经营者

EU AI Act将参与人工智能价值链的主体进行了缜密划分。Provider是指开发人工智能系统或通用人工智能模型，并以自己名义推向市场的人。无论是直接开发，还是委托开发后以自身商标推出，均属Provider。Deployer是指在职业语境下，在自己权限内使用人工智能系统的人。如果医院引入人工智能医疗影像判读仪对患者进行诊疗，该医院便是Deployer。此外，还分别定义了进口商（Importer）、分销商（Distributor）、授权代理人（Authorized Representative）以及产品制造商（Product Manufacturer）。法条中至少出现了7种角色。

韩国《人工智能基本法》仅使用两个范畴。开发经营者（第2条第7项甲目）是指开发并提供人工智能的人。使用经营者（同条乙目）是指利用开发经营者提供的人工智能，提供产品或服务的人。另外还有使用者（第2条第8项）。即接受人工智能产品或服务的人，也就是最终消费者。

在实务上，决定性的差异存在于Deployer和使用经营者之间。在欧盟，医院若使用人工智能医疗影像判读仪，即为Deployer。其必须实施基本权利影响评估，履行透明度义务，并具备人为监督体系。在韩国，如果同一医院使用相同的设备，若仅用于自身的诊疗业务结构中，则依然属于“使用者”，不附带规制义务。但若被评估为利用开发经营者提供的人工智能向患者提供医疗服务的结构，则可能符合使用经营者（第2条第7项乙目）的定义，此时便会伴随透明度义务和高影响经营者的责任。

这种差异构成了韩国经营者进军欧盟市场的实质性障碍。在韩国，引入人工智能的机构的合规负担极低。但在欧盟，引入人工智能的机构作为Deployer也要承担相当大的义务。韩国人工智能企业若想向欧洲的医院或学校提供系统，必须一并提供技术文件、使用说明书和监控工具，以便该医院或学校能够履行Deployer的义务。一直只遵守韩国法律的企业一旦敲开欧盟市场的大门，就会突然面临必须建立Deployer支援体系的状况。

还有一点。欧盟对Provider的定义中包含了“以自己名义或商标（under its own name or trademark）”推出的人这一表述。拿来外部模型并以自己名义提供服务的白标经营者也是Provider。韩国法律中没有与之对应的明文规定。将外部模型进行微调后作为新服务提供的人，究竟是开发经营者还是使用经营者，这还是一个尚未有科学技术信息通信部正式解释的领域。

4. 行政罚款的千倍差距

让我们把数字并列来看。

EU AI Act的最大罚款金额为3500万欧元或全球年营业额的7%，取两者中较高者。按2026年7月的汇率计算，约合610亿韩元。对违反高风险人工智能义务的行为，将处以1500万欧元或营业额3%的罚款。对违反禁止惯例的行为将适用最高水平的制裁。

韩国《人工智能基本法》的最高行政罚款为3000万韩元。行政罚款在三种情况下征收（第43条第1款）：未履行人工智能使用事实的事先告知（第31条第1款）、未指定韩国境内代理人（第36条第1款）、不履行中止命令或纠正命令（第40条第3款）。未遵守文档化义务的情况并不会直接成为行政罚款的对象，只有在收到纠正命令后仍不履行时，才会导致行政罚款。

约610亿韩元与3000万韩元。这是超过千倍的差距。韩国的3000万韩元，在法律施行前就备受批评，被认为难以赋予年营业额达数千亿韩元的大企业以守法动机。在科学技术信息广播通信委员会审议过程中，当行政罚款标准引发争议时，国内人工智能产业正处于初期成长阶段、过度的处罚会萎缩创新等逻辑，成为了低金额的依据。若算上辅导期，在2027年1月之前，对遗漏透明度义务甚至不会处以行政罚款。

欧盟的高额罚款并不能保证绝对有效。自GDPR施行以来，虽然处以了数千万欧元级别的罚款，但企业提出法律异议并进行数年争讼的情况很多。草率得出韩国低额行政罚款必然无效的结论也是不可取的。相比于行政罚款，声誉损失或被排除在政府采购之外，可能会成为更强烈的守法诱因。

尽管如此，超过千倍的差距用数值展现了两部法律的意志差异。欧盟在法文中铭刻了将规制违规行为制裁至影响企业存续水平的意志。而韩国则选择了相比处罚更倾向于通过市场诱因来引导合规的战略。

5. 禁止惯例：韩国所没有的

EU AI Act第5条从源头上禁止了具体的人工智能惯例。有必要浏览一下整个禁止清单。

社会评分（social scoring）被禁止。无论是公共机构还是民间组织，都不允许利用人工智能将社会行为转化为分数并给予不利影响。潜意识操纵（subliminal manipulation）也被禁止。以人们无法意识到的方式引导行为的人工智能、滥用弱点的人工智能都在此列。执法领域的实时远程生物识别原则上被禁止，仅承认

防止恐怖主义等极其有限的例外。工作场所和教育机构的情感识别（emotion recognition）人工智能也包含在禁止范围内。医疗或安全目的的情况则有例外。通过2026年的数字一揽子法案修订，未经同意生成私密图像的人工智能和生成儿童性剥削物的人工智能被迫加进清单（自2026年12月2日起适用）。

韩国《人工智能基本法》中没有此类禁止范畴。不存在通过法律从源头阻断特定人工智能类型的条款。它仅仅赋予高影响人工智能经营者以风险管理和透明度义务，并未宣布完全不能开发或运营特定用途的人工智能。

这一空白究竟是韩国法律的弱点还是灵活性，因立场而异。欧盟公民社会认为禁止条款是保护基本权利的最低底线。其逻辑是，允许社会评分的社会与禁止社会评分的社会有根本性不同。韩国产业界部分人士则反驳称，在技术快速演进的情况下，如果事先禁止用途，可能会连带着阻断那些无法预测的有益运用。

从实务角度看，有一个重要的事实。韩国经营者若想在欧盟提供服务，必须遵守欧盟的全部禁止清单。不能因为韩国法律中没有禁止规定，韩国企业就可以在欧盟市场运营社会评分人工智能。第5条适用于所有向欧盟区域内提供结果的Provider和Deployer。

6. 通用人工智能（GPAI） vs 生成式人工智能

EU AI Act引入了通用人工智能（General-Purpose AI, GPAI）的概念。GPAI模型（第3条第63项）是指“利用大量数据通过大规模自监督学习训练而成的，表现出相当的通用性并能执行广泛任务的”模型。GPAI系统（第3条第66项）则是基于该模型可用于各种目的的系统。在此之上还追加了“系统性风险（systemic risk）”这一上位范畴。如果学习累计运算量超过 10^{25} 次方FLOP，就会被推断为存在系统性风险的GPAI，并被赋予更严格的义务（风险评估、事故报告、确保网络安全）。

韩国法律中没有与GPAI对应的概念。取而代之的是定义了“生成式人工智能”（第2条第5项）。即“模仿输入数据的结构和特性，生成文字、声音、图像、影

像等结果的人工智能系统”。此外，法律第32条和施行令第24条规定了确保安全性义务的对象。这通常被称为“高性能人工智能”范畴。要求学习累计运算量在 10^{26} 次方FLOP以上，应用了当前最前沿人工智能技术，并且风险程度可能对人的生命、身体安全及基本权利产生广泛且重大的影响。必须同时满足这三个条件才是适用系统。

差异有两点。在运算量标准上，欧盟是 10^{25} 次方，韩国是 10^{26} 次方。相差十倍。如果是欧盟标准属于规制对象的模型，按韩国标准则可能被排除在对象之外。截至2026年7月，在韩国国内开发的人工智能模型中，超过 10^{26} 次方FLOP的几乎没有。这实质上起到了缩小规制对象的效果。

概念的广度也不同。欧盟的GPAI并不局限于内容生成。只要具备分类、推理、工具使用、代码编写等“通用能力”的模型，就是GPAI。韩国的“生成式人工智能”则聚焦于“生成”结果的系统。不生成内容但具备通用推理能力的人工智能模型是否被韩国法律的生成式人工智能定义所涵盖，目前尚不明确。

7. 欧盟人工智能办公室 vs 国家人工智能战略委员会

统筹EU AI Act执行的机构是欧洲人工智能办公室（European AI Office）。其于2024年2月21日在欧盟委员会内部成立，拥有125名以上员工，由6个科室组成。即人工智能与机器人卓越科、规制与合规科、人工智能安全科、人工智能创新与政策协调科、社会公益科、卫生与生命科学科。该办公室拥有对GPAI模型的直接执法权，并且通过数字一揽子法案修订，还追加了对集成在大型在线平台（VLOP/VLOSE）中的人工智能系统的监管权限。各个成员国还分别设有国家监管机构（National Competent Authority），负责执行高风险人工智能系统的市场监督。

韩国的治理结构有所不同。国家人工智能战略委员会是最高级别的政策协调机构。委员长由总统担任，委员名额不超过60人，下设8个分科委员会进行运营。通过2026年修订法，其政策协调和成果管理权限得到了扩大。实务层面的执行由

科学技术信息通信部负责。被指定为人工智能政策中心的韩国智能信息社会振兴院（NIA）负责政策开发和基本计划支援，人工智能安全研究所（ETRI附属）负责安全评估和安全技术研究。

差异在于执行力。欧盟人工智能办公室拥有对GPAI Provider的直接调查权限和罚款征收权限。一旦确认违规，可以自行下达纠正命令并施加制裁。韩国的战略委员会只是政策协调机构，而非执行机构。事实调查和行政罚款的征收由科学技术信息通信部负责，但在辅导期间，优先采取的是行政指导和纠正建议。尽管48个中央行政机关的CAIO（首席人工智能官）协商会每月召开，但这只是政府内部的协调机制，而非针对经营者的执法机构。

8. 韩国经营者在欧盟市场需要额外遵守的事项

一直只遵守韩国法律的经营者如果进军欧盟市场，就必须额外履行韩国法律中没有的义务。让我们来梳理一下具体的清单。

首先是检查禁止惯例。必须确认自家的人工智能是否落入欧盟第5条的禁止清单。由于韩国法律中没有禁止范畴，在韩国合法的人工智能功能在欧盟可能会被禁止。在招聘过程中分析应聘者情感的人工智能、在学校中识别学生情感状态的人工智能就是典型案例。

必须重新进行高风险人工智能分类。欧盟的高风险人工智能判定由双重路径（产品安全路径和附件III路径）与豁免机制（第6条第3款）构成。韩国的11个高影响领域与欧盟的附件III领域在很大程度上重合，但并不完全一致。欧盟列举了生物识别、核心基础设施、教育、雇佣、核心服务获取、执法、移民管理、司法、民主程序。韩国法律中没有关于移民管理或民主程序的单项规定。

必须通过符合性评估。在韩国，自主影响评估就已足够，但在欧盟，必须具备技术文件、质量管理体系、符合性声明，并获得CE标志，才能进入市场。在某些情况下，还需要第三方认证。

必须建立Deployer支援体系。在韩国，引入人工智能的机构大部分作为“使用者”处于规制之外，但在欧盟，同样的机构则是Deployer。韩国人工智能企业向欧洲客户提供系统时，必须提供使用说明书、日志记录体系、人为监督工具，以便该客户能够履行Deployer义务。

必须确认GPAI义务。即使不属于韩国的“生成式人工智能”或“高性能人工智能”范畴，只要按欧盟标准符合GPAI定义，就会产生技术文档化、公开学习数据摘要、遵守版权政策的义务。如果符合系统性风险GPAI，还会追加风险评估、事故报告以及网络安全措施。

必须指定授权代理人。在欧盟没有营业场所的Provider必须在欧盟区域内指定授权代理人（Authorized Representative）。这与韩国法律的韩国境内代理人制度类似，但欧盟代理人的角色和责任范围更广。

透明度义务的水平不同。韩国法律要求对人工智能运作事实进行事先告知以及标示生成物标识。欧盟在此基础上，还要求分别公开人工智能系统正在与人互动的事实、情感识别系统正在运行的事实以及属于深度伪造内容的事实（第50条）。

还需要注意时间限制。通过数字一揽子法案修订，高风险人工智能义务的适用时间对于附件III的独立型被推迟至2027年12月2日，对于附件I的产品内置型被推迟至2028年8月2日。这段宽限期虽然争取了准备时间，但禁止惯例（2025年2月施行）、GPAI义务（2025年8月施行）、透明度义务（2026年8月施行）已经处于适用中或即将适用。

9. 致站在两部法律之间的经营者

韩国《人工智能基本法》给予经营者广阔的自主权。EU AI Act则向经营者赋予了缜密的义务。不能因为韩国法律宽松就意味着欧盟法律正确，也不能因为欧盟法律严格就意味着韩国法律错误。这两部法律反映了不同的社会共识。韩国判定人工智能产业的增长是当务之急，而欧盟则判定对市民基本权利的保护优先于产业振兴。

留给经营者的问题并非哲学争论，而是实务。如果自家的人工智能服务有哪怕一点点进军欧盟市场的可能性，仅遵守韩国法律是不够的。现在必须检查是否符合欧盟标准的高风险定义、是否落入禁止惯例、能否通过符合性评估，以及需要向Deployer提供哪些资料。

韩国《人工智能基本法》是全球首部全面施行的综合性人工智能法律。EU AI Act高风险义务的适用则要到2027年12月。由于韩国率先起步，韩国经营者有时填补两部法律之间的鸿沟。至于还剩多少时间，取决于其商业模式。就在向欧洲客户发送第一份提案书的那一刻，时间便已经开始流逝。

第25章 2026年欧盟综合修正案对韩国经营者的实务启示

2025年11月19日，欧盟委员会（European Commission）发布了旨在调整《人工智能法案》（AI Act）核心义务适用时间的《数字综合法案》（COM(2025) 836 final）。该法案的核心内容是推迟刚刚公布1年零4个月的法律的核心义务适用时间。2026年5月，欧洲议会和欧盟理事会达成临时协议，并由成员国予以确认，目前仅剩在官方公报上刊登。高风险人工智能的监管时间表将发生全面改变。对于已进入或正准备进入欧盟市场的韩国经营者而言，这既是重新调整应对战略的机遇，也是风险。

1. 综合修正案的背景，为何如此迅速地进行修改？

欧盟《人工智能法案》（Regulation 2024/1689）于2024年8月1日生效。该法案原计划自2026年8月2日起适用大部分条款。然而，该法案在生效后约15个月便出台了修正案。

其背景是欧洲人工智能产业界的强烈反对。法国的Mistral、德国的Aleph Alpha、芬兰的Silo AI等欧洲人工智能企业抱怨称，合格评定费用负担及监管不确定性导致了投资萎缩。其逻辑是，在欧洲以外的竞争对手（如美国的OpenAI、中国的百度等）在相对宽松的监管下飞速发展的同时，却只有欧洲企业受到了掣肘。

2024年12月正式就任的欧盟委员会主席乌尔苏拉·冯德莱恩（Ursula von der Leyen）的第二届委员会将减轻监管负担作为核心议程。综合修正案正是该议程的具体产物。

2. 被推迟的义务，哪些事项被延后？

修正案的核心是按类型推迟了对高风险人工智能系统（High-risk AI systems）义务的适用时间。

独立型（standalone）高风险人工智能系统的适用时间从2026年8月2日推迟至2027年12月2日。即额外增加了约16个月的宽限期。附件三（Annex III）中所列领域（生物识别、关键基础设施、教育、就业、执法等）的人工智能系统即属于此类。

内置于产品安全法律（product safety legislation embedded）的人工智能系统的适用时间从2027年8月2日推迟至2028年8月2日。这适用于医疗器械、汽车、航空、机械等受现有产品安全规定约束的产品中包含人工智能的情形。该类型属于现有产品认证体系和《人工智能法案》合格评定双重适用的结构，推迟的决定反映了两个体系之间的协调需要更多时间。

对于韩国经营者而言，此项延期的效果是显而易见的。准备进入欧盟市场的高风险人工智能经营者在合格评定的准备上额外获得了16个月的时间。出口医疗器械人工智能的经营者则获得了到2028年8月的宽限期。

然而，延期并非豁免。在延期期间，欧盟人工智能办公室（AI Office）仍将发布针对高风险人工智能系统的指南，并推进标准化工作（harmonised standards）。上述日期属于“最迟期限”，因此如果欧盟委员会确认已具备标准等合规支持手段，则可以提前适用。不可将延期视为休息时间，而应当将其用作准备时间。

3. 未被推迟的义务，2026年8月2日将生效哪些内容？

并非所有事项都被推迟。既有已经实施的义务，也有按原定日程于2026年8月2日适用的义务。

通用人工智能（GPAI）模型义务（第51条至第55条）已于2025年8月2日起实施。GPAI模型提供者负有编写技术文档、制定遵守著作权法的政策、公开关于训练数据的足够详细的摘要的义务。存在系统性风险的GPAI模型还被附加了红队测试、网络安全措施及重大事故报告义务。

透明度义务（第50条）将按计划于2026年8月2日实施。即公开人工智能系统与人类进行交互的事实（聊天机器人事先告知）、生成物标识（包括深度伪造标签）、人工智能生成的文本涉及公共利益事项时的标识义务。但根据综合协议，针对现有系统的机器可读标识（水印）要求获得了至2026年12月2日的4个月额外宽限期。

对GPAI的执法权也将于2026年8月2日启动。欧盟委员会（及其下属的AI Office）将能够直接监督GPAI模型提供者，并在其违规时处以行政罚款（全球营业额的3%或1500万欧元中的较高者）。

在韩国经营者中，向欧盟市场提供聊天机器人或生成式人工智能服务的企业，必须在2026年8月2日前准备好履行透明度义务。这一时间点与韩国《人工智能基本法》修正案的施行时间（2026年7月21日）仅相差12天。如果建立起同时履行这两部法律透明度义务的体系，则可以减少双重工作。

4. 与韩国法律施行时间点的比较时间表

将这两部法律的施行时间点并列来看，可以发现一种模式。

2025年2月2日，欧盟《人工智能法案》的禁止性人工智能实践（第5条）开始适用。已在实施中。

2025年8月2日，欧盟《人工智能法案》的通用人工智能（GPAI）模型义务开始适用。GPAI技术文档、著作权政策、训练数据摘要义务自此生效。

2026年1月22日，韩国《人工智能基本法》第一阶段施行。基本义务（事先告知、生成物标识）以及高影响、高性能人工智能义务也将生效。但预计在施行初期，将以引导为主进行管理。

2026年7月21日，韩国修正案施行。扩大可及性、促进投资等修改和新设的条款将生效。

2026年8月2日，欧盟《人工智能法案》透明度义务（第50条）开始适用并启动GPAI行政罚款权限。即聊天机器人事先告知、深度伪造标签。

2027年以后，韩国的引导为主管理结束（预计）。正式的执法将开始。

2027年12月2日，欧盟独立型高风险人工智能义务适用（综合修正案后的时间点）。

2028年8月2日，欧盟内置于产品的高风险人工智能义务适用。

如果韩国经营者同时进军欧盟市场，则可以制定这样一份路线图：在2026年下半年履行双方的透明度义务，并最迟从2027年上半年起着手准备履行欧盟的高风险义务。

5. 韩国经营者针对欧盟市场的实务路线图

具体而言，应该在何时做什么？

立即：如果是GPAI模型提供者，应检查技术文档的编写、著作权政策的制定以及训练数据摘要的公开状态。鉴于该义务已于2025年8月2日起适用，如果尚未具备，则必须立即履行。

2026年上半年：了解欧盟《人工智能法案》第50条的透明度义务。将与韩国法律规定的事先告知及生成物标识义务相重叠的部分和额外要求整理成对照表。设计同时满足双方要求的告知文案、标签方式和技术措施。

2026年下半年：将聊天机器人事先告知和深度伪造标签应用于针对欧盟市场的服务中。

2027年上半年：追踪欧盟协调标准（harmonised standards）的发布情况。当欧洲标准化委员会和欧洲电工标准化委员会（CEN-CENELEC）正在开发的人工智能质量管理标准、风险管理标准确定后，对本公司产品的合格性进行自我评估。必须在此之前判断独立型高风险人工智能系统是否属于附件三规定的领域，

如果属于，则开始准备合格评定程序（第三方认证或自我合格声明）。

2027年下半年：如果需要第三方认证，应向欧盟境内的公告机构（Notified Body）询问认证日程，并正式开展合格评定。鉴于公告机构的处理能力有限，必须及早行动。

2028年上半年：如果包含内置于产品的人工智能，应完成现有产品认证（CE标志、MDR、机械规定）与《人工智能法案》合格评定的综合文档。

6. 欧盟监管沙盒与韩国监管沙盒

欧盟《人工智能法案》第57条规定了人工智能监管沙盒（regulatory sandbox）。该制度允许成员国在受控环境中测试人工智能系统，并可事先确认其是否符合监管要求。

韩国的监管沙盒是基于《信息通信融合法》（ICT特别法）第37条规定的临时许可及基于第38条之2规定的实证特例制度。《人工智能基本法》并未新设独立的沙盒制度，但在第19条第3款中规定，针对人工智能融合产品及服务，政府应积极提供支持，以使该临时许可和实证特例得以顺利实施。实际上，现有的ICT监管沙盒正在受理与人工智能相关的申请。

两种制度之间存在差异。欧盟沙盒的重点在于事先测试《人工智能法案》的合格性。虽然参与沙盒不能替代合格评定，但如果接受监督机构的指导并获得包含参与结果的书面证明（结项报告），则能够使正式进入市场的准备工作变得更加容易。韩国的监管沙盒则是一种暂时免除现有法令监管以测试业务的机制。

当韩国经营者准备进军欧盟市场时，可以考虑参与欧盟成员国的人工智能监管沙盒。法国、西班牙、荷兰等国正在运营或准备运营沙盒。参与沙盒本身并没有国籍限制，但需要与相关成员国的市场监督机构（Market Surveillance Authority）进行沟通。

7. 减轻SME负担及其对韩国中小企业的启示

综合修正案的另一主轴是减轻中小企业（SME）的负担。

《人工智能法案》原文中也包含与SME相关的条款，但业界不断指出，合格评定费用及编写文件的负担对中小企业而言过于沉重。业界估算，每件高风险人工智能的合格评定可能耗费数万欧元，如果算上建立整个合规体系，成本可能高达数十万欧元。对于员工不足50人的人工智能初创企业而言，这是一笔难以承受的金额。

综合修正案包含了放宽对SME合格评定程序的方案。尽管具体内容将在欧洲议会和欧盟理事会的立法过程中确定，但目前正在讨论扩大自我合格声明（self-assessment）的范围、采用简化的技术文档格式以及减免相关手续费用等措施。

这对韩国中小企业的启示有两点。计划进军欧盟市场的韩国人工智能中小企业应确认其是否能够享受SME减免优惠。符合欧盟法律关于SME定义（员工少于250人，且年营业额不超过5000万欧元或资产总额不超过4300万欧元）的韩国企业可以成为减免对象。即使未在欧盟境内设立法人实体，只要根据《人工智能法案》的域外适用条款受到监管，同样可以享受SME减免的很大一部分优惠。但也有类似优先接入成员国沙盒这种要求在欧盟境内设立营业场所的优惠，因此必须针对每种减免类型确认其具体条件。

与韩国法律的对比也具有一定意义。韩国《人工智能基本法》并未针对高影响、高性能人工智能的义务明文规定按企业规模实行差异化对待。3000万韩元的行政处罚上限实际上同等适用于中小企业和大型企业。未来，韩国法律中也可能出现按企业规模实行差异化义务的讨论。

8. 中国的人工智能监管时间表与三国比较

中国并未像欧盟那样制定单一的人工智能法。取而代之的是，中国陆续出台了各个领域的规定。

《互联网信息服务算法推荐管理规定》于2022年3月施行，深度合成管理规定（《互联网信息服务深度合成管理规定》）于2023年1月施行，《生成式人工智能服务管理暂行办法》于2023年8月施行，《人工智能生成合成内容标识办法》于2025年9月施行。综合性的《人工智能法》尚停留在学界公布专家建议稿及在政府立法计划中被列为研究推进项目的阶段，尚未出台政府层面的正式草案及制定时间。

对比这三个国家/地区的施行时间点，可以发现一个有趣的模式。中国自2022年起已经开始适用监管。韩国将从2026年1月起实施。欧盟将于2026年8月适用透明度义务，并于2027年12月（综合修正案后）适用高风险义务。

仅就施行时间点而言，韩国比中国晚了4年，但比欧盟的高风险监管规定提前了近2年。然而，法律的涵盖范围和执法力度却不尽相同。中国的监管属于分领域规定，因此存在空白；欧盟的规定虽全面，但施行较晚；而韩国则介于二者之间。

在全球市场运营人工智能服务的韩国经营者必须将这三个法域的监管日程整合到一张日历上。必须能够一目了然地看到哪些义务将在哪个国家于何时适用。

9. 应对双重监管的文档单一化战略

需要同时遵守韩国法律和欧盟法律的经营者容易陷入文档双重编写的陷阱。如果将根据韩国法律制定、保存并在要求确认时出示的风险管理方案，与向欧盟市场监督机构提交的技术文档分开制作，则管理成本将翻倍。

解决方案是在一开始便将文档结构设计为同时满足双方要求。

以风险管理文档为例。将韩国法律第34条第1款第1项（风险管理方案的制定和运营）的要求与欧盟《人工智能法案》附件四（技术文档清单）的要求整理成对照表。将韩国要求的事项（风险识别、风险缓解措施、监控体系）与欧盟要求的事项（风险管理系统说明、训练·验证·测试数据规范、性能指标）进行映射。将重叠的事项合并为一个章节，并将韩国专用事项和欧盟专用事项分别作为单独

的附录列出。

这是一种从单一主文档中分别提取韩国提交用版本和欧盟提交用版本的结构。只需维护主文档，便可以在一处反映双方的变更事项。

该战略非常适合医疗器械人工智能经营者。尽管需要四种文档，即韩国食品药品安全处的医疗器械风险管理报告（基于ISO 14971）、韩国《人工智能基本法》的风险管理方案、欧盟MDR的技术文档以及欧盟《人工智能法案》的合格性文档，但其根源均为同一个风险管理系统。

金融人工智能经营者也是如此。韩国金融委员会的人工智能指南、《人工智能基本法》的高影响人工智能义务、欧盟《数字运营弹性法案》（DORA）以及欧盟《人工智能法案》的高风险人工智能义务。将这四层监管文档纳入统一的治理框架进行管理，可以有效减少重复劳动。

文档单一化并非技术问题，而是组织问题。法务部门、开发部门和质量管理部门必须以同一份文档为标准开展工作。如果各部门各自制作符合其管辖规定的文档，则会导致内容脱节和更新滞后。需要确定主文档的所有者，管理变更历史，并建立定期反映双方相关法律变更事项的流程。

欧盟的高风险人工智能义务因综合修正案而被推迟的当下，正是设计文档单一化的最佳时机。与其在义务生效后仓促制作，不如利用延期期间搭好框架，这样在成本和质量上都更具优势。

第26章 美国科罗拉多州、犹他州法律与NIST AI RMF：全球经营者的准据点

2026年初，一家在科罗拉多州丹佛市拥有客户的韩国人力资源（HR）科技企业收到了一封电子邮件：“请发送确认贵公司AI招聘面试分析服务是否属于《科罗拉多州人工智能法》（CAIA）规定的高风险AI系统的文件。如果属于，请一并提交旨在防止算法歧视的影响评估报告和数据治理文件。”当时，该法的生效日期定于2026年6月30日。首尔总部的法务团队并不知道该法的存在。他们知道美国联邦层面没有全面的AI规制法。但他们认为既然没有联邦法律，规制就会很宽松，这种猜测是错误的。仅在2025年一年，美国各州议会就通过了150项与AI相关的法律。科罗拉多州的法律在此之前的2024年就已制定，而招聘被包含在该法规定的八个核心领域之内。然而，就在法务团队准备合规文件的2026年5月，科罗拉多州议会在生效日期到来之前废除了该法，并以新法取而代之。在美国州法的世界里，规则本身就是这样变动的。

1. 没有联邦法律的国家的50个州实验

截至2026年7月，美国尚未制定全面的联邦AI法律。欧盟已经实施了《AI Act》，韩国实施了《人工智能基本法》，中国实施了《生成式人工智能服务管理暂行办法》。作为世界最大的AI产业国，却没有与这些法律相对应的单一联邦法律。

联邦层面并非完全没有行动。2023年10月30日，拜登总统签署了第14110号行政命令《安全、可靠、可信的人工智能开发与使用（Safe, Secure and Trustworthy Development and Use of Artificial Intelligence）》。该文件指示联邦机构制定AI安全指南、确立红队评估标准、开发深度伪造检测与标识工具，并报告AI对劳动力市场的影响。由于它是行政命令而非法律，除了对部分大规模基础模型开发者和计算提供者要求报告义务外，并未直接对普通民间经营者施加义务。尽管如此，这也是联邦政府设定AI规制方向的信号。

这种信号并未持续太久。2025年1月20日，特朗普总统在上任第一天撤销了第14110号行政命令。拜登政府的AI监督框架在一天之内消失了。同年12月，特朗普总统签署了方向截然相反的行政命令《确保国家人工智能政策框架（Ensuring a National Policy Framework for Artificial Intelligence）》。该命令指示在司法部设立“AI诉讼工作组”，在联邦法院从法律上挑战过度的州AI法律，并命令商务部识别给企业造成过度负担的州法。其中甚至包含了将部分联邦预算与州政府克制AI立法相挂钩的条款。

联邦政府先占（preemption）州法的尝试在议会受挫。在众议院阶段，2025年《预算调整法案》中包含了一项在10年内阻止州AI法律执行的暂停条款。参议院在2025年7月1日的表决中以99票对1票删除了该条款。共和党 and 民主党均反对联邦政府剥夺州政府的AI立法权限。

结果是，美国的AI规制由州议会主导。根据斯坦福大学HAI的《2026年AI指数报告》，从2016年到2025年，美国州议会通过的AI相关法律累计数量为：加利福尼亚州62项、马里兰州28项、弗吉尼亚州25项、犹他州24项、纽约州18项、得克萨斯州17项。只有密苏里州和罗得岛州没有通过任何AI相关法律。其余48个州正以各自的速度和方向推进规制AI的庞大立法实验。

2. 《科罗拉多州人工智能法》：对抗算法歧视的首部全面州法

《科罗拉多州人工智能法》（Colorado Artificial Intelligence Act, CAIA）的法案编号为SB 24-205。该法于2024年5月17日由州长贾里德·波利斯（Jared Polis）签署，原定生效日期为2026年2月1日。然而，该法未能实施哪怕一天。

波利斯州长在签署的同时公开了一封表达担忧的信函。他警告称，该法的合规义务可能会给AI初创企业和中小企业带来负担，从而阻碍创新。他要求议会在生效日期前重新审查该法，并进行缩小适用范围的修订。在2025年8月的特别会议上，科罗拉多州议会没有修改该法的内容，而是通过了一项将生效日期推迟到2026年6月30日的法律（SB 25B-004）。而在推迟后的生效日期到来之前的2026

年5月14日，波利斯州长签署了废除CAIA并代之以新法的SB 26-189。新法以对自动化决策技术（ADMT）的透明度和信息披露为中心进行设计，核心义务从2027年1月1日起适用。防止算法歧视这一沉重的义务体系一次也未施行，便在两年后消失了。

尽管已被废除，CAIA的设计依然留存下来，成为此后美国州法讨论的起点。该法的核心概念是“重大决定（consequential decision）”。法律列举了八个领域：教育入学和机会提供、就业和招聘、金融和贷款服务、基本公共服务、医疗服务、住房、保险、法律服务。在这些领域，如果AI做出了在法律上产生重大或类似影响的决定，或者AI成为该决定的实质性因素（substantial factor），则将该系统视为“高风险AI系统（high-risk AI system）”。

“算法歧视（algorithmic discrimination）”的定义也存在于法律文本中。这指的是基于年龄、肤色、残疾、民族、性别、是否服兵役等受保护特征的不利待遇。韩国《人工智能基本法》中没有直接与之对应的概念。韩国法律建议通过基本权利影响评估来自主检查歧视风险，但并未单独定义或禁止算法歧视。

CAIA区分了开发者（developer）和部署者（deployer）。在科罗拉多州开展业务并开发或实质性修改高风险AI系统的人是开发者，部署并使用该系统的人是部署者。这一结构对应于韩国法律中的开发经营者和使用经营者。

首先看开发者的义务。开发者必须向部署者提供说明高风险AI系统可预见风险和用途的文件。该文件必须包含系统性能评估方法、防止算法歧视措施、数据治理措施。数据治理措施是指数据来源的适当性、潜在偏见及偏见缓解措施。预期的输出类型和系统监控方法也必须包含在文件中。如果开发者发现系统可能引起算法歧视，必须在90天内通知州总检察长（Attorney General）和相关部署者。

部署者的义务则面向消费者。必须向消费者事先告知使用了AI系统。这是即使对非高风险AI也适用的基本义务。只有在理性人能够明显感知到AI介入的情况下才能豁免。当高风险AI对消费者做出不利决定时，附加了额外义务。必须说明决

定的理由、AI的贡献程度、处理过的数据类型和来源。必须向消费者提供纠正错误个人信息的机会，以及对不利决定提出异议并要求人工复核的机会。每年必须实施影响评估，对系统进行重大修改时，需要在90天内进行额外的影响评估。

针对小型部署者有豁免条款。如果员工不足50人，不使用自有数据训练AI，仅按照开发者的预期用途使用，并向消费者提供开发者的影响评估，即可免除包括影响评估义务在内的CAIA的大部分义务。但基本的AI使用告知义务不予豁免。

科罗拉多州总检察长拥有独占的执行权。不认可个人直接起诉企业的私人诉权（private right of action）。举证责任在于经营者一方。开发者和部署者必须直接证明其遵守了CAIA的要求。

法律也设计了防御手段。如果通过内部检查、红队测试或外部反馈发现违法事实并予以纠正，这可以作为积极抗辩（affirmative defense）。正在遵守NIST AI风险管理框架（AI RMF）或ISO/IEC 42001、或者总检察长指定的风险管理框架的情况，也可以成为积极抗辩的依据。正是这一条款，成为将NIST AI RMF提升为美国AI规制事实标准的通道。

3. 犹他州《AI政策法》：告知义务的起跑线

犹他州州长于2024年3月签署了《AI政策法》（Artificial Intelligence Policy Act, SB 149）。该法于2024年5月1日生效，使犹他州成为美国最早实施全面AI规制的州之一。

与科罗拉多州选择禁止算法歧视这一沉重义务不同，犹他州选择了轻松的开局。核心义务只有一个。使用生成式人工智能与消费者互动的个人或企业，在消费者提出要求时，必须明确告知其正在与AI对话。并非在所有互动中都必须自动告知，而是采取消费者询问时才回答的方式。但在2025年的修订中，在使用生成式人工智能提供医疗等需要许可证的受规制职业服务时，告知义务得到加强，要求在消费者询问前事先告知。

犹他州没有建立新的规制机构，而是在现有的消费者保护体系内进行执行。取而代之的是，在商务部下属设立了“人工智能政策办公室（Office of Artificial Intelligence Policy）”，负责运营研究AI应用的“AI学习实验室（AI Learning Laboratory）”项目。这是一种理解优先于规制的路径。

2025年3月，犹他州议会通过了一项单独的法案HB 452。这是对心理健康聊天机器人的规制。当AI提供情感咨询或心理健康支持服务时，增加了三项义务：必须向使用者告知其正在与AI对话；不得将使用者的个人健康数据出售或共享给第三方；即使刊登广告，也必须明确说明是广告并披露赞助商，且不得将使用者输入的对话内容用于定向广告。加利福尼亚州也在同年10月制定了SB 243，规定伴侣聊天机器人运营者有义务告知AI身份和自杀相关的安全协议。这是心理健康AI聊天机器人正在成为美国州法新规制对象的信号。

与韩国《人工智能基本法》相比，犹他州的路径规制强度要低得多。韩国法律规定，对使用高影响人工智能或生成式人工智能的产品和服务，事先告知是强制性的（第31条第1款），并对高影响AI施加了风险管理、说明、人工干预、文档化等义务。而在犹他州，甚至只有在消费者要求时才需要告知。然而，在美国国内，犹他州的法律被评价为“先做起来再说”的务实模式。这是因为它与试图一次性规制所有事项却导致任何法案都无法通过的联邦议会形成了鲜明对比。

4. NIST AI风险管理框架：虽为自愿但不可忽视的标准

美国国家标准与技术研究院（NIST）于2023年1月26日发布了AI风险管理框架（AI Risk Management Framework, AI RMF 1.0）。它不是法律。它是不具法律约束力的自愿性指南。然而，该文件在美国AI规制生态系统中所占的分量，远比“自愿”一词所暗示的要重得多。

AI RMF提出了衡量AI系统可信度的七项标准。它们是：有效性与可靠性（Validity and Reliability）、安全性（Safety）、安全与弹性（Security and Resilience）、责任与透明度（Accountability and Transparency）、可解释性

与可解读性（Explainability and Interpretability）、隐私保护（Privacy）、公平性（Fairness）。这七项标准与韩国《人工智能基本法》的基本原则（第3条）或欧盟《AI Act》的要求高度重合。这意味着AI规制的国际通用语言已经形成。

该框架的执行结构由四个核心功能（core function）组成。

治理（Govern）涵盖全局。它是在整个组织范围内建立和维护AI风险管理治理结构的功能，涉及政策、程序、责任分配、组织文化。要使其余三个功能在现场发挥作用，必须首先确立治理。韩国《人工智能基本法》要求经营者制定和运营风险管理方案（第34条第1款第1项），并要求公布管理和监督高影响人工智能人员的姓名和联系方式（施行令第27条第1款），这正是与该功能相对应的。

映射（Map）旨在把握背景。通过分析部署AI系统的环境和条件来识别并对风险进行分类。它会追问：哪些人受到系统的影响？在哪些条件下系统可能失效？偏见可能从哪里引入？韩国法律中判断高影响AI的程序，即确认领域适用性及风险影响和重大性的结构（法律第33条、施行令第25条），与此功能十分相似。

衡量（Measure）利用定量指标和定性评估来分析并追踪已识别的风险。持续监控风险的规模、频率及系统性能的变化。韩国法律要求高影响AI经营者定期完善风险管理方案，即对应于此功能。

管理（Manage）负责分配资源并进行应对。基于映射和衡量的结果，计划并执行旨在降低风险的具体措施。决定是接受、转移、缓解还是规避风险。

这四个功能不是循环一次就结束的检查表，而是贯穿AI系统整个生命周期的可重复适用的循环结构。

AI RMF之所以不再仅仅停留在自愿指南的层面，有一个决定性的契机。那就是《科罗拉多州人工智能法》。CAIA被设计为赋予遵守NIST AI RMF的经营者积极抗辩资格。当被指控违法时，这是一种可以说“我们是按照NIST AI RMF运营风险管理体系的”这一法律防御手段。虽然该法本身在实施前被废除，但它留下了

一个自愿框架可通过州法转化为事实上的法律标准的案例。

NIST在2024年补充发布了生成式AI配置文件（AI RMF Generative AI Profile），补充了针对大型语言模型和多模态AI固有风险的管理方法。其对象包括幻觉（hallucination）、侵犯版权、生成有害内容、CBRN（化学、生物、放射性、核）风险以及信息安全威胁。如果说AI RMF正文是通用AI风险管理的骨架，那么生成式AI配置文件就是在该骨架上编织的生成式AI专有之网。

5. 联邦软规范与州硬规范的共存

美国的AI规制版图呈现出独特的双重结构。

在联邦层面，NIST AI RMF、行政命令和各机构指南作为软规范（soft law）存在。它们要么没有法律约束力，要么像行政命令一样随着总统更迭可能被撤销。拜登政府的AI行政命令在上任第一天就消失的案例便是明证。只要联邦议会无法通过全面的AI法律，联邦层面的AI规范就始终是不稳定的。

在州层面，法律（statute）作为硬规范（hard law）发挥作用。科罗拉多州CAIA的最终历程展现了其属性。虽然联邦司法部介入了违宪诉讼，联邦法院甚至做出了停止执行的决定，但实际废除该法的最终还是科罗拉多州议会本身。原则上，州法的修订和废除掌握在州议会手中。伊利诺伊州的《生物特征信息隐私法》（BIPA）自2008年施行以来，已连续18年有效。州法的延续与总统更迭无关。

这两层架构交汇的地带，就是CAIA的积极抗辩条款。其结构是，如果遵守联邦的自愿框架（NIST AI RMF），即可作为对违反州法的法律防御。这就将软规范纳入了硬规范的合规标准之中。未来，随着越多的州法参考NIST AI RMF，这一自愿框架就越接近于事实上的全国标准。尽管它并非法律，却成为了难以不遵守的基准。

特朗普政府2025年12月的行政命令加剧了这一结构的紧张。司法部AI诉讼工作组宣布，将在联邦法院对“负担过重”的州法发起挑战。然而，正如参议院99对

1的表决结果所显示的，两党均反对联邦政府夺走州政府的AI立法权限。尽管州儿童安全、数据中心基础设施、州政府采购被提及为仍留有州规制余地的领域，但其余领域的边界并不明晰。这种紧张状态仍未消除。

欧盟和韩国的规制体系从根本上不同于这种双重结构。欧盟《AI Act》是直接适用于27个成员国的条例（Regulation）。各成员国制定相互冲突法律的空间受到限制。韩国《人工智能基本法》在全国范围内统一适用。无论在首尔还是在济州，规则都是一样的。只有美国实行各州不同的AI规制体系。

6. 韩国经营者进军美国：解读各州规制地图

当韩国AI企业进军美国市场时，所面临的规制环境因州而异。同一产品每次跨越州界，都会碰壁于不同的规则。

截至2026年7月，各州情况可分为三类：正在施行、即将施行、废除与替代。若要在犹他州运营AI聊天机器人服务，现在就必须在消费者要求时披露AI的身份，如果是精神健康聊天机器人，则禁止向第三方出售或共享健康数据，并禁止基于使用者输入内容投放定向广告。也有即将施行法律的州。加利福尼亚州《AI透明度法》（SB 942）要求从2026年8月2日起，一定规模以上的生成式人工智能提供者必须提供水印和免费检测工具；华盛顿州的HB 1170要求从2027年2月1日起，月活跃使用者达到100万人以上的生成式人工智能系统提供者必须提供AI生成媒体的出处数据（provenance data）。此外，还有废除或替代法律的州。原打算要求AI招聘工具每年进行防止算法歧视影响评估的科罗拉多州CAIA在实施前就消失了，从2027年1月1日起，将适用以透明度和信息披露为中心的新法。

伊利诺伊州需要特别注意。自2008年起施行的《生物特征信息隐私法》（BIPA）虽然不是AI专用的法律，但直接适用于面部识别AI。如果在没有获得个人事先书面同意的情况下收集生物特征信息，个人可以直接对企业提起诉讼。Meta因涉嫌在Facebook照片标签功能中未经同意收集面部识别数据，同意支付6.5亿美元（约合8,500亿韩元）达成和解。Clearview AI在2025年交出公司23%股份（估值

约5,175万美元)的和解方案曾获得法院批准,但在2026年7月的上诉审中,因缺乏代表当事人要件而被撤销批准,案件发回一审。与科罗拉多州CAIA仅将执行权交给总检察长的结构不同,在伊利诺伊州个人可以提起集体诉讼。诉讼风险的维度完全不同。

得克萨斯州的情况则表现出不同的一面。《负责任的AI治理法》(Responsible AI Governance Act, HB 149)于2025年通过并于2026年1月实施,但相较于原法案,大幅减少了私营部门义务,重点聚焦于禁止行为操纵、生成儿童性剥削材料、故意歧视和侵犯宪法权利。同样是在“AI规制”的名义下,各州规制的广度和深度截然不同。

与韩国的规制环境相对比,差异非常鲜明。韩国只有一部《人工智能基本法》适用于全国。从首尔到釜山都适用同一套规则。若想在美國做生意,就必须逐一确认服务所覆盖的所有州的法律。

现实的应对之策有两条路径。一条是以规制水平最高州的基准作为通用基础。只是各州法律在适用对象、告知时点、健康信息与生物特征信息及广告规范上互不相同,用一项严格的标准并不能自动满足所有州的要求,因此必须单独确认各州的特别规定。另一条路径是绘制各州的规制地图,并根据各州的要求调整服务。这需要更多成本,但能够扩大业务范围。

无论采取哪条路径,以NIST AI RMF为基准构建风险管理体系都是基础。因为被废除的科罗拉多州法曾有过将该框架引入作为积极抗辩标准的先例,未来其他州法也有可能做出类似的借鉴。

7. 韩国、欧盟、美国：三种规制哲学的比较

这三个体系从不同的问题出发。

欧盟《AI Act》问道：“这个AI系统是否有侵犯基本权利的风险？”如果风险无法接受,则予以禁止。如果是高风险,必须经过事先符合性评估才能进入市场。

它是直接适用于27个成员国的单一条例，对违反最严重禁止行为的罚款最高可达3,500万欧元或全球年营业额的7%。禁止进行社会评分和潜意识操纵，除搜寻失踪人员等狭窄的例外情况外，执法目的的公共场所实时远程生物特征识别也被禁止。其规制明确且强力。因此，经营者的事先负担也很重。

韩国《人工智能基本法》问道：“这个AI系统对社会有重大影响吗？”它对高影响AI施加了风险管理和透明度义务，但没有事先符合性评估或禁止范畴。它采用自主努力义务和公共采购激励这种软规制。其条件是：全国单一适用、最高行政罚款3,000万韩元，以及政府宣布的一年以上的辅导期。这是一种将促进创新和保护权利融于一法的混合路径。

美国在联邦层面本身就没有统一的问题。科罗拉多州曾问“这个AI存在歧视吗？”，但在替换法律时将问题改为了“消费者知道这属于自动化决定吗？”；犹他州问“消费者知道它是AI吗？”；伊利诺伊州问“获得生物特征信息的同意了吗？”。NIST AI RMF问的是“具备风险管理体系了吗？”，但它没有法律约束力。50个州正在各自提出问题并给出答案。

以数字来衡量执行力。欧盟的最高罚款额为3,500万欧元，按2026年汇率约合595亿韩元。中国仅凭《个人信息保护法》，即可处以最高5,000万元人民币（约100亿韩元）或上一年度营业额5%两者中金额较高者的罚款。韩国的最高罚款为3,000万韩元。美国则因州而异。科罗拉多州是以总检察长为中心的制裁，伊利诺伊州BIPA则允许行使私人诉权提起集体诉讼，可能产生数亿美元的和解金，其实质风险可能远超法定上限。

在这一比较中有一点是清晰的。在基本权利保护的系统性方面，欧盟在三者中处于领先地位。在产业发展速度和经营者负担的平衡方面，韩国的软规制占据优势。从实验和多样性的角度来看，美国各州分散的模式是一个同时测试50种不同路径的巨大实验室。无法断言这三种模式中哪一种是正确的。只是对韩国经营者而言，唯一确定的是，他们必须面对读懂所有三种规制环境并为之做好准备的现

实。

如果韩国AI经营者同时进军美国和欧盟，必须运营三重合规体系：具备符合欧盟事先符合性评估水平的技术文件和风险管理体系；在美国，以NIST AI RMF为基准满足各州的附加要求；在韩国，履行《人工智能基本法》的透明度义务和高影响AI判断程序。由于欧盟标准的规制水平在三者中最高，以遵守欧盟规定为基准线，补充韩国和美国差异的模式是比较现实的。

虽然这种三重结构是一种负担，但对韩国经营者来说并非完全不利。因为韩国《人工智能基本法》在很大程度上与欧盟《AI Act》和NIST AI RMF共享核心原理。透明度、风险管理、人工监督、公平性这四大支柱是三个体系所共有的。如果是忠实履行韩国法律的经营者，就已经具备了适应美国和欧盟要求的出发基础。剩下的工作就是准确测量从出发基础到各市场目的地之间的距离，并填补其中的空白。

第九部 行业别经营者实务指南

第27章 金融人工智能经营者

2023年秋，在首尔汝矣岛的一家储蓄银行，一名30多岁个体户的贷款申请被拒绝了。窗口职员无法解释原因，只能向其展示信用分数未达到标准的画面。打出该分数的不是人，而是人工智能信用评估模型。输入了哪些数据，根据什么标准扣减了分数，职员不知道，银行总行也无法给出明确答复。根据2020年修订的《信用信息法》，该个体户有权要求对自动化评估结果进行说明，但他甚至没有被告知自己享有这项权利。

金融人工智能与其他行业的人工智能性质不同。能否获得贷款、能否投保、投资资金应投向何处，都由机器来决定或推荐。一旦出错，一个人的整个经济生活都会动摇。因此，金融人工智能经营者不仅要遵守《人工智能基本法》，还要同时受金融委员会单设指南和《信用信息法》的规制。这种被称为双重规制的结构，给金融人工智能经营者带来了怎样的负担，又设置了怎样的保护机制，我们将逐一展开分析。

1. 双重规制的结构：金融委员会指南与《人工智能基本法》

金融领域的人工智能规制有两层。一边是金融委员会于2021年7月首次出台、2026年6月修订的《金融领域人工智能指南》，另一边是2026年1月22日施行的《人工智能基本法》。

金融委员会的指南不是法律。这是行政机关向金融公司和金融科技企业建议的方针。虽然没有法律约束力，但在金融监督院的检查中，是否遵守该指南实际上发挥着评估项目的作用。对金融公司而言，这具有无法无视建议的现实约束力。

《人工智能基本法》是法律。无论是金融领域、医疗领域还是教育领域，只要是人工智能经营者，均适用该法。第5条第1款规定，“除其他法律有特别规定外，依照本法规定执行”。如果《信用信息法》中已经有人工智能相关规定，则优先适用该规定；如果没有，则适用《人工智能基本法》。金融委员会指南不是法

律，因此不具备相对于《人工智能基本法》的特别法地位。

在实务中理顺这种结构是这样的。运营人工智能贷款审查系统的银行必须根据《信用信息法》第36条之2的规定，向使用者说明自动化评估的主要标准和基础信息概况。同时，根据《人工智能基本法》第2条第4项G目的规定，该系统可能属于高影响人工智能，如果属于，则根据第34条第1款的规定，会增加制定风险管理方案、说明判断依据、制定使用者保护方案、确保人为干预体系、文件保存5年的义务。在此基础上，金融委员会指南还要求具备治理体系、遵循辅助手段性原则并确保可靠性。

2. 金融委员会人工智能指南的核心原则

金融委员会指南在2021年7月首次出台时，提出了金融公司引入和运营人工智能时应遵守的五项原则。2026年6月18日发布并于6月22日起施行的修订版将其重编为七项原则。分别是治理、合法性、辅助手段性、可靠性、金融稳定性、诚实信用和安全性。前三项原则从一开始就延续下来，而此前的可解释性和公平性虽然从独立原则的名称中去除了，但其要求已融入新的原则体系中。

首先是治理。这是要求建立针对人工智能系统开发与运营的内部控制体系。应在董事会或高管团队层面上批准人工智能应用战略，并指定负责管理人工智能相关风险的责任人。这相当于在CISO（首席信息安全官）或CRO（首席风险官）的角色中增加了人工智能风险监管的职责。这对中小金融科技企业来说是一种负担。对于没有余力配备专职人员的初创企业而言，如何满足建立董事会层面治理体系的要求，是一个现实课题。

其次是合法性。这是指在应用人工智能时遵守相关法律法规的原则。这听起来理所当然，但在金融领域，适用法律法规的范围很广。《信用信息法》、《电子金融交易法》、《资本市场法》、《保险业法》和《个人信息保护法》相互交织。运行一个人工智能模型，必须同时检查这五部法律的要求事项。

第三是辅助手段性。这是指人工智能的判断不应是最终决策，而应是协助人类判断的辅助手段的原则。即使人工智能拒绝了贷款审批，也必须维持由人类审查该结果并做出最终决定的结构。这与《人工智能基本法》第34条第1款第4项所要求的人为管理和监督方向一致。实务中的问题在于形式上的人为干预。如果人工智能给出“拒绝”的结论，职员只是不按下批准按钮而已，实际上推翻判断的情况很少。要使辅助手段性原则真正发挥作用，必须同时制定职员能够推翻人工智能结果的权限和标准。

第四是可靠性。这是指必须能够信任并使用人工智能判断结果的原则，而可解释性是其核心。必须能够向使用者解释人工智能为何得出这样的结果。《信用信息法》第36条之2是将这一要求以法律形式体现的条款。可以要求说明自动化评估的结果、主要标准和基础信息概况。在基础信息不准确或过时的情况下，可以要求更正或删除，或者要求重新计算结果。金融委员会指南比法律更进一步，建议努力对人工智能模型的判断逻辑本身进行说明。也就是说，即使使用黑箱模型，也要应用事后（post-hoc）可解释的技术。

对公平性检查的要求也随之而来。必须检查人工智能，以防其因性别、年龄、种族、出生地区等因素得出歧视性结果。信用评估模型系统性地给特定地区居民打低分，或者保费计算模型根据性别造成不当差异，都是违反公平性的典型案例。如果训练数据中融入了过去的歧视模式，人工智能就会复制这种模式。指南要求检查训练数据的偏见，并监控结果的公平性。

修订版确立的新原则是金融稳定性、诚实信用和安全性。这包含了这样的问题意识：人工智能的故障不仅会损害个别客户的利益，还可能动摇整个金融市场的稳定；确认了金融公司即使使用人工智能，也依然承担对客户的诚实信用义务；并包含了防范针对人工智能系统安全威胁的要求。

3. 人工智能信用评估与贷款审查

人工智能信用评估是金融人工智能中历史悠久的应用案例。NICE评估信息公司、韩国信用局（KCB）等个人信用评估公司（CB）运营着通过输入数百个变量来计算个人信用分数的人工智能模型。银行、信用卡公司和储蓄银行将该分数用于贷款审查。

《信用信息法》直接规制这一领域。第36条之2规定了对自动化评估的说明及提出异议的权利。“自动化评估”是指信用信息公司等从业人员不参与评估业务，仅通过计算机等信息处理装置进行处理的评估。根据该条文，个人信用信息主体可以提出四项要求：是否进行自动化评估；自动化评估的结果；主要标准；基础信息的概况。

在此基础上，又增加了《人工智能基本法》。第2条第4项G目“招聘、贷款审查等对个人的权利与义务关系产生重大影响的判断或评估”中明确规定了贷款审查。人工智能贷款审查系统属于高影响人工智能。根据第34条第1款的规定，高影响人工智能经营者被赋予了制定和运营风险管理方案（第1项）、制定和实施说明方案（第2项）、制定和运营使用者保护方案（第3项）、人的管理与监督（第4项）以及制作和保存履行依据文件（第5项）的义务。在主页上发布主要内容和文件保存5年等细节事项由施行令第27条予以规定。

NICE评估信息公司等CB公司的定位很复杂。NICE开发人工智能信用评估模型并提供给金融公司。在《人工智能基本法》上，它是开发经营者。同时，NICE本身作为《信用信息法》规定的个人信用评估公司，直接承担自动化评估说明义务。金融公司（银行、信用卡公司）接收NICE的模型并将其应用于贷款审查，因此在《人工智能基本法》上是使用经营者。如果开发经营者NICE履行了风险管理方案、说明方案和使用使用者保护方案，只要使用经营者银行没有实质性地改变模型的功能，这三项义务就视为已履行（施行令第27条第3款）。但是，人的管理与监督以及文件的制作与保存，银行必须单独履行。这是防止双重负担的机制。

避免重复规制在《人工智能基本法》施行令中也有规定。施行令附表1按项目规定了遵守其他法令的哪些义务，即可视为履行了《人工智能基本法》的哪些责任。如果遵守《信用信息法》第35条之2的说明义务，并具备该法第36条之2第1款规定的说明程序，则认定为履行了《人工智能基本法》第34条第1款第2项规定的说明方案义务。但是，附表1中没有的义务，例如制定风险管理方案或保存文件5年，则必须根据《人工智能基本法》单独履行。

4. 智能投顾：当人工智能推荐投资时

智能投顾是指通过算法分析投资者的倾向和目标，构建投资组合并自动进行再平衡的服务。在韩国，从2016年起，金融委员会就运营了智能投顾沙盒。在非面对面全权委托投资等部分业务模式中，通过Koscom运营的沙盒实际上发挥着必要条件的作用。

《资本市场法》对投资咨询业和全权委托投资业进行规制。如果智能投顾“推荐”投资，则属于投资咨询；如果直接管理客户委托的资金，则属于全权委托投资。在全权委托投资的情况下，必须进行符合不同客户风险倾向的管理，并充分告知发生损失的可能性。不能因为是人工智能在管理，就免除这项义务。

与《人工智能基本法》的契合点如下。关键在于智能投顾决定或推荐个人资产配置的行为，是否属于第2条第4项G目“对个人的权利与义务关系产生重大影响的判断”。它不像贷款审查那样明确列在法律条文中。其是否包含在“等”字之内，尚有解释的空间。如果投资推荐出错，会产生本金损失这种重大的财产影响。它被分类为高影响人工智能的可能性很高。不过，截至2026年7月，尚未确认科学技术信息通信部有正式的权威解释。

如果金融委员会指南的辅助手段性原则适用于智能投顾，那么即使人工智能构建了投资组合，最终的买卖决定也需要人为确认的程序。完全自动化的全权委托投资（通过算法自动执行买入和卖出的方式）与辅助手段性原则之间存在张力。金融委员会尚未明确消除这一张力。

5. 人工智能保险审查

在保险业中，人工智能主要用于两个方面：核保审查（underwriting）和保险金索赔审查（claims processing）。

在核保审查中，人工智能通过分析保险申请人的健康信息、职业、生活习惯和过去事故记录，来决定是否承保以及保费多少。传统上由保险精算师完成的很大一部分工作正由人工智能替代。在保险金索赔审查中，人工智能可辨别索赔文件的真伪，并将欺诈的可能性进行量化打分。

《保险业法》规制保险产品的批准和条款审查，但没有直接规制人工智能核保审查本身的条款。金融委员会指南填补了这一空白。根据公平性原则，必须检查人工智能核保审查模型是否对特定疾病或遗传信息施加了不当的权重。

在与《人工智能基本法》的关系中，关于保险核保审查是否属于高影响人工智能，存在不同解释。第2条第4项G目写有“招聘、贷款审查等对个人的权利与义务关系产生重大影响的判断或评估”。拒绝投保对个人的经济权利有重大影响。认为其包含在“等”字之中的解释更为自然。保险金索赔审查也是如此。拒绝支付保险金的判断直接影响个人的财产权。

在保险业中，人工智能公平性问题之所以成为争议焦点，是因为训练数据的特性。保险本质上是一项对风险进行分类的业务。对高风险群体收取高额保费是保险的原理。然而，人工智能可能会以“风险分类”的名义，实际上实施歧视。对特定邮编地区居民系统性地收取高额保费的模型，会成为地域歧视的工具。保险业中风险分类与禁止歧视之间的界限，在人工智能出现之前就已存在争议，但人工智能使这一界限变得更加模糊。

6. 作为高影响人工智能的金融人工智能：第2条第4项G目

《人工智能基本法》第2条第4项列举了高影响人工智能的11个适用领域。G目是“招聘、贷款审查等对个人的权利与义务关系产生重大影响的判断或评估”。

将该条文代入金融人工智能中，贷款审查是明确包含在内的。招聘也包含在内。“等”字扩大了范围。信用评估、保险核保审查、保险金支付审查、投资适当性评估等，都可能属于“对个人的权利与义务关系产生重大影响的判断”。

一旦被分类为高影响人工智能，根据第34条第1款的规定，经营者将被赋予五项义务。

必须制定并运营风险管理方案。必须具备政策制定、风险识别、分析与评估、处理、改善的循环体系。由于金融公司已经具备了操作风险管理体系，将人工智能风险纳入现有体系是现实的方法。

必须制定并实施说明方案。在技术可行的范围内，必须能够说明人工智能得出的最终结果、用于判断的主要标准以及训练数据的概况。根据施行令第27条第1款的规定，必须在营业场所或主页上发布风险管理方案的主要内容、说明方案的主要内容、使用者保护方案以及管理监督者的姓名和联系方式，属于商业秘密的事项可以剔除。

必须制定并运营使用者保护方案。这是要求设立预防和应对使用者在使用人工智能产品或服务过程中可能受到损害的机制。

必须具备人可以干预的体系。需要由人来审查人工智能的决定，并在必要时能够推翻的程序。这与金融委员会指南的辅助手段性原则方向一致。

必须将履行依据文件保存5年（施行令第27条第2款）。这与金融公司的记录保存义务（通常为5年至10年）的期限有重合。

7. 双重规制负担与防止重复的机制

金融人工智能经营者感受到的负担是这样的。运营一个人工智能贷款审查系统，会叠加《信用信息法》的自动化评估说明义务、《人工智能基本法》的高影响人工智能义务、金融委员会指南的七大原则以及《个人信息保护法》的自动化决定相关规定。实务中的核心问题是：优先履行哪部法律的哪个条款，就能被视为

满足了其他法律的义务。

《人工智能基本法》设有避免重复规制的机制。第5条第1款规定，“除其他法律有特别规定外，依照本法规定执行”。如果《信用信息法》中有人工智能相关的规定，则《信用信息法》优先。

施行令的方向也一样。施行令第27条第5款和附表1，按项目规定了如果遵守《个人信息保护法》、《数字医疗产品法》、《信用信息法》等现有法令的哪些义务，即可视为履行了《人工智能基本法》的哪些责任。

具体来说，相互认可的义务总结如下。

说明义务是相互认可的。如果遵守《信用信息法》第35条之2的说明义务，并具备该法第36条之2第1款规定的说明程序，则无需另外履行《人工智能基本法》第34条第1款第2项的说明方案义务（附表1第2项C目）。

使用者保护方案也是相互认可的。如果履行了《金融消费者保护法》第10条规定的所有责任，则视为履行了《人工智能基本法》第34条第1款第3项的使用者保护方案（附表1第3项B目）。

提出异议及要求纠正的权利情况有所不同。《信用信息法》第36条之2第2款规定了要求纠正及删除基础信息权、要求重新计算权，其与《人工智能基本法》要求人工管理及监督的宗旨相重叠。然而，附件1并未将其规定为视为履行项目，因此不能仅凭宗旨重叠而免除该义务。

也有无法相互认可的领域。附件1采取了限定列举认可项目的方式，关于贷款审查，并没有认可建立风险管理方案、人工管理及监督、文件保存5年的项目。包括在主页上发布等在内，这些义务必须另行履行。

金融委员会指南不是法律，因此不属于与《人工智能基本法》相互认可的对象。无论多么忠实地履行指南，也不能免除《人工智能基本法》的义务。相反，即使

履行了《人工智能基本法》，如果无视指南的建议，也可能在金融监督院检查中受到指责。

8. 《信用信息法》第36条之2：自动化评估的说明与提出异议

《信用信息法》第36条之2是金融人工智能规制的实质核心。其在2020年修订数据三法时新设，自2020年8月5日起施行。

该条文的适用对象即“自动化评估”有两种：个人信用评估（信用评分计算）以及是否设定、维持金融交易及决定内容的评估（贷款审批、额度设定等）。适用对象的经营者为个人信用评估公司（CB）和总统令规定的金融公司。根据施行令第31条之2，属于《关于设立金融委员会等法律》第38条各项规定的人员，即银行、保险公司、证券公司等接受金融监督院检查的所有金融机构，均符合此条件。

要求说明的范围很广。可以要求说明是否进行自动化评估、评估结果、主要标准以及基础信息的概述这四个方面。

还有提出异议的权利。可以提交对自己有利的信息，如果基础信息不准确，可以要求纠正或删除，或者要求重新计算结果。要求重新计算权在实质上意味着可以对人工智能的决定进行重新审查。

拒绝理由仅限于三种情况：本法或其他法律有特别规定，或为了遵守法令义务而不可避免的情况；难以设定及维持金融交易等商业交易关系的情况；总统令规定的类似情况。由于拒绝理由是限定列举的，因此若不符合这三种情况，便不能拒绝说明。

与《人工智能基本法》的说明义务（第34条第1款第2项）相比，《信用信息法》更加具体且强制性更强。《人工智能基本法》第34条第1款第2项规定要“在技术可能范围内”制定说明方案，赋予了经营者较大的自由裁量权。而《信用信息法》第36条之2明文规定了必须说明的项目（结果、主要标准、基础信息概述）

，并限定列举了拒绝理由。对于金融人工智能经营者而言，《信用信息法》的要求构成了其实质标准。

9. EU AI Act中金融人工智能的地位

欧盟人工智能法（EU AI Act）明确将金融人工智能分类为高风险（high-risk）。

附件III（Annex III）第5项涵盖了“获取和享受基本私人服务及基本公共服务和福利”（Access to and enjoyment of essential private services and essential public services and benefits）领域。其中包含了自然人的信用度评估（creditworthiness evaluation）和信用评分决定（credit scoring）。人寿保险和健康保险的风险评估及定价（risk assessment and pricing in relation to natural persons in the case of life and health insurance）也被分类为高风险人工智能。

韩国关注影响，而欧盟关注风险。这一差异在金融人工智能中产生了具体的不同结果。若要在欧盟推出人工智能信用评估系统，必须通过事先符合性评估（conformity assessment）。必须编制技术文件，构建风险管理系统，具备数据治理能力，并制定人工监督措施。必须在欧盟数据库中进行注册。不过，将这一高风险义务的适用时间从原定的2026年8月推迟至2027年12月的程序实际上已处于收尾阶段。如有违反，将处以最高1500万欧元或全球销售额3%中的较高金额作为罚款。

在韩国，相同的人工智能信用评估系统即使被分类为高影响人工智能，也无需接受事先符合性评估。只需自主建立并运行风险管理方案即可。即使违反了义务，也不会立即被处以行政罚款，而是在接到科学技术信息通信部的纠正命令后仍未履行时，才会被处以最高3000万韩元的行政罚款（第40条第3款、第43条第1款）。欧盟与韩国在规制强度上的差异，在金融人工智能领域表现得淋漓尽致。

EU AI Act在保险领域也非常具体。在人寿保险和健康保险的风险评估中使用人工智能，会被分类为高风险。汽车保险和财产保险则被排除在外。这意味着只有

直接涉及人的生命和健康的保险才被视为高风险。韩国《人工智能基本法》没有按保险种类进行区分。按照“对个人的权利与义务关系产生重大影响的判断”这一概括性标准，不论是何种保险种类，均可能属于高影响人工智能。

10. 金融人工智能初创企业的合规实务

汝矣岛和江南的金融科技企业面临的现实情况如下。

规制的数量很多。包括《人工智能基本法》、《信用信息法》、《个人信息保护法》、《资本市场法》（投资相关）、《电子金融交易法》（支付相关）、《保险业法》（保险相关）。根据经营模式，可能会同时适用三四部法律。此外，还会叠加金融委员会指南的要求。

对初创企业也有有利的一面。预告称《人工智能基本法》的宣导期至少为1年，如有需要还可以延长。在宣导期内，虽然也会产生法律义务本身，但对于遗漏透明性义务或文件化不完备的情况，行政指导和纠正建议将优先于行政罚款等制裁措施生效。可以在人工智能基本法支持台接受匿名咨询。规模越小的初创企业，利用这一宣导期逐步完善合规体系就越符合实际情况。

开发经营者和使用经营者的角色分担也减轻了负担。开发人工智能模型的金融科技企业作为开发经营者，如果履行了建立风险管理方案、说明方案以及使用者保护方案等义务，那么引进并使用该模型的金融公司（使用经营者）只要不实质性改变模型功能，就无需另外履行这三项义务。人工管理及监督、文件制定及保存义务则需由各使用经营者自行履行。对于金融科技企业而言，自身作为开发经营者已完成合规这一点，将成为其对金融公司的业务竞争力。

最困难的一点是，法律和指南的解释尚未稳定确立。针对“对个人的权利与义务关系产生重大影响的判断”是否包含资产管理推荐、保险费计算、海外汇款风险评估等情形，科学技术信息通信部尚未作出权威解释。在《人工智能基本法》上，进行事实调查和下达纠正命令的权限在科学技术信息通信部（第40条），但如何与负责监督金融公司的金融委员会及金融监督院进行角色分工并展开协商，

目前处于未理清的状态。

11. 金融人工智能规制的空白

可以看出存在两处空白。

《信用信息法》第36条之2以“个人信用评估”以及“是否设定、维持金融交易及决定内容”为对象。并未明确规定保险核保审查和保险理赔审查。《保险业法》中也没有关于人工智能审查的直接规定。保险领域的人工智能处于法律的空白地带中，仅靠金融委员会指南在松散地填补这一空白。

对于在高频交易（HFT）和算法交易中使用人工智能的情况，相关规则也有所欠缺。《资本市场法》虽然对算法交易有一定的规定，但对于人工智能对市场造成的系统性风险（如闪崩等事件）并没有全面的规范。EU AI Act也未将该领域分类为高风险。这是因为金融市场的稳定性是与个人权利和义务受到侵害不同层面的问题。然而，正如2010年美国的闪崩（Flash Crash）事件和2012年骑士资本（Knight Capital）事件所表明的那样，人工智能算法交易的故障可能会动摇整个市场。

12. 金融人工智能经营者现在应做的事

运营金融人工智能的经营者，其检查清单如下。

确认是否已具备《信用信息法》第36条之2规定的自动化评估说明体系。必须要有在使用者要求说明时，能够提供评估结果、主要标准、基础信息概述的程序和系统。这是金融人工智能合规的起点。

判断自身的人工智能系统是否属于《人工智能基本法》第2条第4项G目的高影响人工智能。如果用于贷款审查，则属于此列。如果用于信用评估，则极有可能属于此列。如果用于保险核保审查或投资适当性评估，也有极高的可能会被认定为此类人工智能。

如果属于高影响人工智能，就要建立风险管理方案。在现有的运营风险管理体系中追加人工智能风险项目的作法更为高效。要建立起将履行依据文件保存5年的体系。

检查金融委员会指南的七大原则。即治理、合法性、辅助手段性、可靠性、金融稳定性、诚实信用、安全性。对于各项原则，评估目前本公司的履行水平，并对不足之处加以完善。

在法律稳定实施之前，应先确保与监管部门的沟通渠道畅通。人工智能基本法支持台的匿名咨询是免费的。金融监督院的金融科技支持中心也接受咨询。在解释尚不确定的领域，事前主动询问优于事后遭受指责。

金融人工智能规制是按照两套“时钟”运行的。《信用信息法》的时钟已经运转了6年。《人工智能基本法》的时钟则始于2026年1月22日。在两个时钟指向同一方向的地方，可以通过一次履行同时满足两部法律的要求。在方向不同的地方，则必须分别应对。事先掌握哪个时钟指向哪个方向，是金融人工智能经营者迈出合规的第一步。

第28章. 医疗及生物人工智能经营者

这里是首尔某大学医院的影像医学科。2026年2月的某个周一早晨，放射科专科医生打开了胸部CT影像。在屏幕右侧，AI读片辅助软件正用黄色圆圈标出三处疑似肺结节的部位。该软件已获得食品医药品安全处（以下简称“食药处”）颁发的3类医疗器械许可。同时，根据半个月前施行的《人工智能基本法》，用于保健医疗提供及使用体系的AI系统可能属于高影响人工智能。专科医生的脑海中浮现出一个务实的问题：既然已经获得了食药处的许可，是否还需要单独履行《人工智能基本法》规定的义务？在这部法律中，我究竟是经营者还是使用者？

1. 双重监管的结构

医疗AI经营者置于双重监管体系之中。

其一是食药处的医疗器械许可审查体系。医疗器械法令根据危害程度将医疗器械分为1类至4类。搭载AI的软件医疗器械（SaMD, Software as a Medical Device）大多被分类为2类或3类。3类以上必须通过食药处的专门技术文件审查和临床试验资料评估并获得许可后，方可用于患者。仅在2023年一年内，食药处许可的AI医疗器械就有64件，2024年增加到了108件。

其二是《人工智能基本法》对高影响AI的监管。《人工智能基本法》第2条第4项将高影响AI的适用领域列为从（甲）目至（十一）目共11个，其中医疗领域出现了两次。（丙）目是根据《保健医疗基本法》构建和运营保健医疗的提供及使用体系。（丁）目是根据《医疗器械法》开发及使用医疗器械，以及根据《数字化医疗产品法》开发及使用数字化医疗器械。

两项监管发生重叠。获得食药处许可的AI影像读片软件在接受《医疗器械法》规制的同时，也被归类为《人工智能基本法》中的高影响AI。开发经营者既要向食药处提交技术文件和临床数据，又必须履行《人工智能基本法》第34条第1款规定的风险管理、说明、使用者保护、人工管理与监督以及文件化义务。

2. 食药处许可体系：AI医疗器械如何进入市场

食药处对AI医疗器械的审查与传统医疗器械审查有所不同。

传统医疗器械一旦获得许可，只要设计不发生改变，该许可就会一直维持。而AI医疗器械随着训练数据的追加或算法的更新，其性能会发生变化。为了应对这一问题，食药处多次修订了许可审查指南。2025年1月24日，食药处发布了全球首个生成式人工智能医疗器械许可审查指南。

从类别分类的实务来看，大多数AI辅助诊断软件都被归入3类。肺部CT辅助读片、脑部CT出血检测、乳腺X光读片、视网膜影像糖尿病视网膜病变检测软件均属于此类。即使是不代替医生判断而仅起辅助作用的软件，只要读片结果会对诊断产生影响，就会升至3类。4类是指在没有医生干预的情况下独立进行诊断或做出治疗决定的AI，目前在韩国还没有4类AI医疗器械获得许可的先例。

许可审查周期曾是一个痛点，平均需要390天。在2022年10月引入“创新医疗器械一体化审查评估制度”后，该周期缩短到了80天。对于企业而言，进入市场的速度提高了5倍。

并不是获得了许可就能立即在医院使用，还需要厘定国民健康保险的收费标准。截至2024年7月，已有8款AI产品被指定为“创新医疗技术”，7款产品获得了“新医疗技术评估延期”。被定为选择性给付的收费标准在310韩元至2,920韩元之间。由于单次CT读片仅增加数千韩元的费用，AI医疗器械企业的盈利模式目前仍不稳定。

3. 《数字化医疗产品法》：全球首部专属法律

2024年1月，《数字化医疗产品法》正式公布。这是全球首个专门针对数字化医疗产品制定的独立法律。

该法对三类对象进行了区分：数字化医疗器械是指应用了AI和先进数字技术的医疗器械；数字化融合药品是指将药品与数字化医疗器械或数字化医疗·健康支

持设备相结合的产品；数字化医疗·健康支持设备是指用于医疗支持和健康维持的数字技术产品，不属于医疗器械。

主要条款的施行日期分为两个阶段。本法正文的大部分内容已于2025年1月24日施行，而与数字化医疗·健康支持设备相关的条款（第33条至第35条）则于2026年1月24日施行。

与《人工智能基本法》的纽带恰恰就在这里。《人工智能基本法》第2条第4项（丁）目将根据《医疗器械法》规制的医疗器械与“根据《数字化医疗产品法》规制的数字化医疗器械”并列。在（丁）目中，关于数字化医疗器械部分的施行日期为2026年1月24日。由于定义了数字化医疗器械的《数字化医疗产品法》正文已于2025年1月24日起施行，因此该日期是为了配合《数字化医疗产品法》完成最后阶段条款的全面施行。

梳理两部法律的关系如下：《数字化医疗产品法》由食药处管辖，负责审查产品的安全性和有效性；《人工智能基本法》由科学技术信息通信部管辖，负责管理AI系统的社会影响和可信度。同一款AI医疗器械同时受到这两个部门的监管。

4. 作为高影响人工智能的医疗AI

如果对《人工智能基本法》第2条第4项的（丙）目和（丁）目进行详细解读，医疗AI被归类为高影响人工智能的范围比想象的要广。

（丙）目的“保健医疗提供及使用体系的构建与运营”引用了《保健医疗基本法》第3条第1项。该法将保健医疗定义为“国家、地方自治团体、保健医疗机构或保健医疗人员等为保护和促进国民健康而进行的一切活动”。这不仅包括诊疗室中使用的AI读片系统，还可能涵盖医院的患者分配算法、急诊室重症度分类AI以及传染病预测系统。

（丁）目则更为直接，其对象是根据《医疗器械法》规制的医疗器械以及根据《数字化医疗产品法》规制的数字化医疗器械的“开发及使用”。不仅是开发搭

载AI医疗器械的经营者，利用该器械向患者提供服务的医院，也可能被纳入监管范围之内。

梨花女子大学法学专门大学院的李元福教授明确指出了这一点：“不仅是搭载了人工智能的医疗器械及数字化医疗器械，医疗机构为了使诊疗顺利进行而使用的人工智能系统，都会被自动分类为高影响人工智能。”

一旦被归类为高影响AI，经营者被赋予的义务包含在第34条第1款这一条文五个项中：风险管理方案的制定与运营（第1项）；最终结果、用于得出该结果的主要标准以及训练用数据概要的说明方案的制定与实施（第2项）；使用者保护方案的制定与运营（第3项）；对高影响人工智能的人工管理与监督（第4项）；可确认措施内容的文件的撰写与保管（第5项）。风险管理方案、说明方案以及使用者保护方案的主要内容，管理与监督人员的姓名和联系方式，必须在办公室或官方网站等处进行公示（《施行令》第27条第1款），且有关措施的依据文件必须保管5年（《施行令》第27条第2款）。

获得食药处许可的AI医疗器械开发经营者已在许可过程中提交了临床数据和技术文件。虽然这与《人工智能基本法》规定的文件化义务有重叠之处，但两部法律的要求并不完全一致。食药处的技术文件侧重于产品的安全性和有效性。而《人工智能基本法》的文件化要求则涵盖了对基本权利的影响、偏见管理以及人工干预体系。两者的文件名称不同，管辖机构不同，提交受理部门也不同。

5. 具体案例：从影像读片到新药研发

在AI医疗器械市场中，成熟度较高的领域是影像读片。

Lunit的“Lunit Insight”是可在胸部X光和乳腺造影影像中标记异常特征的AI软件。该软件已获得食药处许可和欧盟CE认证，并在南美和东南亚销售。VUNO的“Vunomed”是用于分析胸部X光、CT、MRI影像的产品系列，旗下多款产品拥有欧盟CE认证，并开发了心跳骤停预测算法。JLK则提供可在脑部MRI中进行早期诊断和病变检测的AI。

这些企业均属于《人工智能基本法》规定的开发经营者。他们负责开发AI并向医院提供。由于在属于高影响AI领域的（丁）目中运营，因此承担着高影响AI经营者的义务。

影像读片之外的领域也在快速增长。

AI新药研发（AI drug discovery）利用机器学习来进行候选物质筛选和分子结构优化。这是指在庞大的化合物库中，筛选出与特定蛋白质靶点结合概率较高的分子的工作。对于此类AI系统是否属于《医疗器械法》所规定的医疗器械，目前还存在争议。如果是并不直接用于患者的研究工具，则可能不属于医疗器械。然而，如果其参与到保健医疗的提供及使用体系中，则仍有可能根据（丙）目被归类为高影响AI。

AI基因组分析将AI应用于基因变异检测、表达模式分类及疾病风险预测。如果基因组分析软件获得了医疗器械许可，则属于（丁）目。尽管仅用于研究目的的分析工具是否适用《医疗器械法》尚不明确，但如果其结果会对临床决策产生影响，则可能会被纳入保健医疗提供体系（即（丙）目）。

Sky Labs的指环型血压计“CART BP”于2023年获得了食药处许可。这是全球首款通过欧洲高血压学会（ESH）验证的无袖带血压测量设备。SELVAS AI的“Selvy Medivoice”基于语音识别自动撰写医疗记录。Life Semantics的“CanopyMD SCAI”可通过智能手机筛查皮肤癌。Medi Whale则通过视网膜影像评估心血管风险。

these 产品共同面临着相同的问题：向食药处提交的文件与《人工智能基本法》所要求的文件，在多大程度上发生重叠，又在何处产生分歧？

6. 义务重叠与消除：第5条的局限性

《人工智能基本法》第5条（与其他法律的关系）是为解决这一问题而制定的通用条款。

第1款规定：“除其他法律对人工智能等另有特别规定的情形外，均适用本法规定。”如果《医疗器械法》或《数字化医疗产品法》中存有与AI相关的“特别规定”，则该规定将优先于《人工智能基本法》。

问题在于“特别规定”的范围非常模糊。《医疗器械法》针对AI医疗器械的安全性和有效性审查设有详细条款。这些条款对于《人工智能基本法》规定的制定风险管理方案的义务（第34条第1款第1项）而言，究竟属于“特别规定”，还是仅为针对其他事项的其他规定，目前尚不明确。

仁荷大学法学专门大学院的郑灿模教授举例指出，在司法实践中，《个人信息保护法》第6条第1款的类似条款未能解决与其他个人信息相关法律之间的冲突。他认为，仅凭《人工智能基本法》第5条的通用条款，很难消除与《医疗器械法》或《数字化医疗产品法》之间的监管重叠。

李元福教授则说得更为直接，他认为第5条的通用条款“在医疗AI领域中，很难期待其能对消除《人工智能基本法》与《医疗器械法》或《数字化医疗产品法》之间的监管重叠起到实质性帮助”。如果其他法律的条款不属于“特别规定”，那么经营者就必须同时遵守这两部法律。

与欧盟相比，欧盟《人工智能法案》为减少与《医疗器械条例》（MDR）的重叠，设置了具体的衔接机制。欧盟《人工智能法案》第43条第3款允许通过现有的公告机构（notified body）进行单一符合性评估。第11条第2款允许以一份技术文件同时满足两部法律的要求。韩国《人工智能基本法》中也存在联动的衔接条款。第34条第3款规定，如果经营者根据其他法令履行了相当于第34条第1款各项的措施，则视为已履行该项措施。《施行令》第27条第5款及附表1具体规定了认定标准。就数字化医疗器械而言，根据《数字化医疗产品法》建立质量管理体系并获得数字化医疗器械软件质量管理标准符合性判定时，视为履行了风险管理方案（第1项）和文件化（第5项）义务；完成制造或进口许可、认证、申报并在软件上显示信息时，视为履行了说明方案（第2项）义务；履行了该法第13条规

定的所有遵守事项时，视为履行了使用者保护方案（第3项）义务；配备质量负责人时，视为履行了人工管理与监督（第4项）义务。然而，该标准仅限于《数字化医疗产品法》规定的数字化医疗器械。对于未经过《数字化医疗产品法》、仅获得《医疗器械法》许可的传统医疗器械，并不适用附表1的视为履行标准，因此第5条通用条款的局限性依然存在。

不过，目前存在一个现实的缓冲因素。《人工智能基本法》规定的行政罚款最高为3,000万韩元，且没有针对经营者违反义务的刑事处罚条款。政府已宣布，计划在法律施行后运营至少为期一年的过渡期。按照这一方针，在此期间将不会处以行政罚款，这相当于给了经营者协调两部法律义务的时间。然而，一旦过渡期结束，义务重叠的问题依然会原封不动地存在。

7. 欧盟：《医疗器械条例》（MDR）与《人工智能法案》的双重监管

在欧盟，医疗AI同样处于两部法律的交汇点，即《医疗器械条例》（MDR, Regulation (EU) 2017/745）与《人工智能法案》（Regulation (EU) 2024/1689）。

欧盟《人工智能法案》将医疗AI归类为高风险的路径有两条。第一条是附件I路径，即根据MDR或《体外诊断医疗器械条例》（IVDR）需要进行第三方符合性评估的医疗器械中包含AI的情形。如果Class IIa及以上的医疗器械中引入了AI，则会被自动分类为高风险AI。第二条是附件III路径，该路径将急救电话的患者分类、急诊重症度分类以及基本公共服务资格审查等特定用途指定为高风险，从而将属于此类用途的医疗AI纳入其中。

在这种双重结构中，会产生四个方面的冲突。

第一，符合性评估程序不同。MDR的认证程序与《人工智能法案》的符合性评估程序（第43条）在文件格式、评估标准及管辖机构上都存在差异。虽然《人工智能法案》第43条第3款允许通过现有的公告机构进行单一评估，但在实务中，MDR的技术文件格式（MDR附件II）与《人工智能法案》的技术文件格式（《人

工智能法案》附件IV) 在结构上是不一样的。

第二，临床试验中存在法律空白。MDR and IVDR明确将试验用医疗器械排除在“投放市场”或“投入服务”的定义之外，这使得许可前的临床试验成为可能。然而，由于《人工智能法案》中没有与之精确对应的除外条款，关于AI医疗器械的临床试验是否属于《人工智能法案》项下的“投入服务”，目前存在着法律释义上的争论。

第三，风险等级被拉高。MDR将Class IIb及以上视为真正的高风险。然而，《人工智能法案》规定，从Class IIa开始，只要需要第三方认证，就会被归类为高风险AI。例如，搭载AI的电子体温计在MDR中属于低风险，但在《人工智能法案》中可能会变成高风险。根据Muehlematter等人于2021年在《柳叶刀-数字健康》(The Lancet Digital Health)上发表的研究，在美国FDA许可的222件AI/ML医疗器械中，仅有3件属于最高风险等级(Class III PMA)，在欧盟的240件中也仅有3件属于Class III。其余绝大多数属于中低风险等级，但在适用《人工智能法案》时可能会被重新分类为高风险。

第四，医生被课以部署者 (Deployer) 义务。

8. 医生究竟是使用者，还是部署者？

在韩国和欧盟，医生的法律地位截然不同。

在欧盟，医生是部署者 (Deployer)。欧盟《人工智能法案》第3条第4款将部署者定义为“在其权限下使用AI系统的自然人或法人”。仅排除个人、非专业活动。如果医生将AI诊断辅助软件用于患者诊疗，该医生 (或医院) 即为部署者。部署者被赋予第26条的义务。即应根据制造商的使用指南使用AI。监测风险。监督运行。保持透明度。保存记录。遵守相关法律和伦理。如果医疗AI属于附件III，且部署者为公共机构或提供公共服务的私营主体，则甚至需要执行根据第27条规定的基本权利影响评估。

在韩国，情况有所不同。韩国《人工智能基本法》第2条第8项将“使用者”定义为“接受人工智能产品或人工智能服务的人”。第2条第7项b目的“使用经营者”是指“利用开发经营者提供的人工智能，提供人工智能产品或人工智能服务的人”。

当医生将AI医疗器械用于患者诊疗时，医生是使用经营者，还是使用者？对此存在分歧。李元福教授分析称，“根据不同的解释，医疗人员或医疗机构可以被视为属于使用经营者，也可以不被视为使用经营者”。如果医院配备AI判读系统并向患者提供影像诊断服务，这属于“利用人工智能提供服务”的行为，因此可被视为使用经营者。相反，如果医生将AI作为诊疗这一本职工作的工具，则也可以将医生视为人工智能的“接受者”，即使用者。

这种模糊性在实务中所产生的影响并不小。

如果是使用者，则免除《人工智能基本法》的经营者义务。既不需要风险管理方案，也不需要保存文件，更不需要在官方网站发布。如果是使用经营者，只要没有重大的功能变更，开发经营者已经履行的风险管理、说明、使用者保护措施，则视为已经履行（施行令第27条第3款）。但是，如果显著变更了原本的目的或用途，则须承担独立义务。

在欧盟，这种区分相对明确。医生是部署者，患者是受影响者。作为部署者的医生，完全适用欧盟《人工智能法案》第26条的义务。有人担忧，这可能会成为一种负担，从而延缓AI在医疗现场的引入。van Kolfshoeten和van Oirschot于2024年在《Health Policy》学术期刊上发表的论文，分析了欧盟《人工智能法案》的部署者义务对医疗AI采用所产生的抑制效应。

韩国法律的模糊性反倒赋予了其灵活性。科学技术信息通信部可以通过指南或权威解释，将医疗机构的地位界定为任何一方。然而，灵活性是以牺牲可预测性为代价的。当医院引入AI医疗器械时，如果无法确定自己承担何种义务，那么推迟引入本身便成为合理的选择。

9. 经营者现在应当确认的事项

AI医疗器械开发经营者的检查事项如下：

首先确认是否获得韩国食品药品安全处许可。如果作为AI医疗器械获得了许可，则说明已经履行了《医疗器械法》的义务。需要从《人工智能基本法》的高影响人工智能义务中，甄别出可以通过向韩国食品药品安全处提交的文件来满足的部分，以及需要额外编制的部分。如果是根据《数字医疗产品法》获得许可、认证、申报，并具备质量管理体系和质量负责人的数字医疗器械，则首先应当对照根据施行令附件1，是否可以被视为已经履行了第34条第1款各项义务的相当一部分。制定风险管理方案（第34条第1款第1项）可能与韩国食品药品安全处的风险管理报告有很大一部分重合，但人的管理与监督体系（第34条第1款第4项）以及基于基本权利视角的影响评估（第35条），很有可能不包含在韩国食品药品安全处的文件中。

作为使用经营者的医院和医疗机构，首先应当明确自身的法律地位。根据是直接购买AI医疗器械用于诊疗，还是向患者提供基于AI的新服务，义务的范围也会有所不同。可以向科学技术信息通信部的《人工智能基本法》援助台进行咨询。

如果是进军欧盟市场的经营者，则应当防范双重监管。韩国食品药品安全处许可、《人工智能基本法》义务、欧盟MDR合格评定、欧盟《人工智能法案》高风险AI义务。四个监管层级叠加在一起。如果从一开始就以满足双方法律要求的格式编制技术文件，则可以减少重复工作。

预计韩国的医疗AI市场年增长率为50.8%（从2023年至2030年）。这一速度快于全球平均水平的41.8%。在快速增长的市场中，监管的不确定性会成为前期投资的绊脚石。经营者应当持续关注，《人工智能基本法》施行令的修订，或者科学技术信息通信部关于医疗AI的指南，是否能够减少这种不确定性。

医疗AI的监管是一个需要穿过两扇门的结构。如果认为只需穿过一扇门即可，就会在另一扇门前停滞不前。必须掌握一份地图，了解这两扇门的钥匙在何处重

合，又在何处分叉。

第29章 平台与生成式AI业务经营者

2026年3月，在首尔江南的一家初创公司办公室里。该公司运营着一项通过调用OpenAI的API自动生成房产房源描述文案的服务。代表在会议室的白板上写下了三个问题：“我们在自己生成的文案中必须加上‘由AI生成’吗？该标识必须是肉眼可见的文字，还是可以隐藏在元数据中？如果OpenAI已经履行了标识义务，我们是不是就可以免除？”《人工智能基本法》第31条所规定的透明度义务，落到每个使用生成式AI的业务模式上时，其分量都有所不同。

1. 根据业务模式的不同，义务的分量也不同

《人工智能基本法》第31条构建了三层透明度义务。

第1款是事先告知。如果拟提供使用高影响人工智能或生成式人工智能的产品或服务，必须事先向使用者告知该产品是基于该人工智能运行的事实。这共同适用于所有生成式AI服务。

第2款是生成物标识。在通过生成式人工智能制作的成果物上，必须进行“这是由AI制作”的标识。无论是文本、图像、视频还是音频，只要是生成式AI的成果物，均须履行这一义务。

第3款是深度伪造特例。在制作并提供难以与真实情况区分的虚拟声音、图像或视频等成果物时，必须以使用者能够明确识别的方式进行告知或标识。这一标准高于第2款。从“应当进行标识”增加附带了“以能够明确识别的方式”这一条件。

在这三层义务中适用哪一项，取决于该经营者提供何种类型的成果物。生成文本的聊天机器人服务与合成人脸的视频生成服务，在同一部法律的同一条款下承担着不同的义务。

2. 聊天机器人服务：事先告知与生成物标识

运营ChatGPT、Claude、Gemini等聊天机器人服务的经营者，适用第31条第1款和第2款。

首先来看第1款的事先告知。施行令第23条第1款列举了四种告知方法：直接记载在产品上或写在服务条款中的方法；显示在使用者屏幕或终端设备上的方法；在提供服务的场所进行张贴的方法；科学技术信息通信部长官认可的方法。如果是基于网页的聊天机器人，只需在对话屏幕的开头添加“本服务基于人工智能运行”的引导语即可。虽然法律条文上允许仅写在服务条款中，但确保透明度指南建议放在使用者容易识别的位置。

第2款的生成物标识有两种方法。施行令第23条第2款将“人能够识别的方法”与“机器能够读取的方法”并列。即使选择机器读取方式，也会附加一个条件：必须通过引导语或语音等方式，向使用者提供1次以上该成果物是由生成式人工智能生成的事实。这意味着仅隐藏在元数据中是不够的。

在实务中，履行该义务的方式因服务而异。ChatGPT在对话屏幕底部常态化显示“ChatGPT can make mistakes”字样。Claude则在服务屏幕和引导语中表明其为AI助手。只要向韩国 market 提供服务，此类标识是否满足《人工智能基本法》第31条第2款的要求，就会成为实务中需要判断的对象。

聊天机器人服务的文本成果物大多不属于第31条第3款的深度伪造特例。因为它们不属于“难以与真实情况区分的虚拟声音、图像或视频”。文本作为一种媒介，其本身在形式上就无法区分真实与虚拟。人写的文章和AI写的文章无法仅凭外观进行区分。第3款所针对的是感官上模仿真实存在的成果物，即照片般的图像、宛如真人的视频、以及宛如真声的声音。

这里存在例外情况。施行令第23条第4款规定，在以下三种情况下免除全部或部分义务：通过服务名称或屏幕可以明显看出基于AI的事实；仅用于经营者内部业务用途；以及科学技术信息通信部长官规定的情况。“ChatGPT”这一服务名称本身就体现了基于AI的事实，因此有可能属于“显而易见性例外”。但是，对

于调用该服务的API来制作其他名称聊天机器人的使用经营者，该例外很可能不适用。

3. 图像生成服务：水印与深度伪造

对于Midjourney、DALL-E、Stable Diffusion等图像生成服务，可能同时适用第31条第2款和第3款。

第2款适用于所有生成式AI成果物。对于AI生成的图像，必须标识这一事实。问题在于方法。可以直接在图像成果物中嵌入“AI Generated”字样。这是人能够识别的方法。也可以根据C2PA等标准，在图像元数据中植入来源信息。这是机器能够读取的方法。即使选择机器读取方式，也必须提供1次以上的引导语。

第3款额外适用的契机是“难以与真实情况区分”的图像。由Midjourney v6以上模型生成的照片级图像就属于此类。例如，与真实照片难以区分的人物图像，或真实再现实际场所的风景图像。在这种情况下，超出了第2款的一般标识义务，必须以使用者“能够明确识别的方式”进行告知或标识。施行令第23条第3款提出了两项具体标准：使用者能够通过视觉、听觉或软件轻松确认内容；以及应当考虑主要使用者的年龄以及身体和社交条件。

通俗地说，这意味着对于看起来像照片的AI图像，应当以肉眼可见的方式添加水印或标签。仅凭元数据很难满足第3款中“能够明确识别的方式”。

DALL-E 3会自动在生成的图像中插入C2PA元数据。据了解，截至2026年，Midjourney尚未强制要求插入元数据。由于Stable Diffusion是开源模型，因此由分发者决定是否加入水印功能。对于在韩国市场运营这些服务或以此为基础提供服务的经营者而言，在机器读取方式之外，同时进行可视化标识是降低法律风险的选择。

第3款后半部分考虑到了艺术性、创意性的表现物。若相关成果物属于艺术性或创意性表现物，或构成其一部分，则可以“不妨碍展览或欣赏等方式”进行告

知或标识。这意味着在展览用AI创作的美术作品时，不是要在作品上重叠打印水印，而是可以在展览指南或说明牌（Caption）上进行标注。

中国的做法更为细致。于2025年9月1日施行的《人工智能生成合成内容标识办法》规定了双重标识义务：显式标识，即使用者肉眼可见的标签；以及隐式标识，即元数据或数字水印。两者均须履行。强制性国家标准GB 45438-2025规定了技术规范。韩国法律允许在两者中选择其一，而中国法律则将两者都规定为强制性义务。

4. 视频生成服务：深度伪造标识的高墙

Sora、Runway、Pika等视频生成服务是第31条第3款的核心适用对象。

视频比图像具有更大的欺骗风险。它会运动、有声音，并且伴随着时间流逝。看起来像真实人物说了其未曾说过的话的视频，或者看起来像发生了实际并未发生的事件的视频。此类成果物完全符合“难以与真实情况区分的虚拟视频”。

第3款所要求的“使用者能够明确识别的方式”在视频媒介中具体意味着什么？根据施行令第23条第3款的标准，使用者必须能够通过视觉或听觉轻松确认。一种方法是在视频角落持续显示“AI生成”水印。另一种方法是在视频的开头和结尾插入“本视频由人工智能生成”的字幕或语音引导。

在此会产生实务上的难题。在视频被分享到社交媒体、下载、编辑以及重新上传的过程中，可视化水印可能会被裁剪或擦除。视频元数据在流经平台时也会被清除。即使最初的经营者履行了标识义务，法律如何处理在流通过程中标识消失的视频最终到达使用者手中的情况，目前还没有明确的答案。

法律第31条第3款的义务主体是“人工智能经营者”。运营视频生成服务的开发经营者，或以该服务为基础运营视频制作服务的使用经营者承担该义务。对于仅仅分享已生成视频的普通使用者，不适用该义务。第2条第8项的使用者并不属于受规范者。

欧盟《人工智能法案》第50条第4款在此问题上更进一步。对于在业务中使用生成深度伪造的AI系统的部署者（deployer），也负有公开该事实的义务。韩国法律仅对经营者施加了义务，而欧盟则将其扩展到了在业务中使用该系统的使用者。

5. 语音与音乐生成服务：合成语音的标识

Suno AI、ElevenLabs、AIVA等服务生成的合成语音和AI音乐，属于第31条第3款中的“声音”。

ElevenLabs可以复制人的声音。如果输入几分钟的语音样本，它就可以用那个人的声音朗读任何文本。如果成果物与真实的人声难以区分，则属于第3款的适用对象。必须以使用者能够明确识别的方式，告知其为合成语音的事实。

声音是无法添加视觉水印的媒介。一种方法是在音频文件的开头或结尾插入“本声音由人工智能合成”的引导语。还有一种技术是在声音信号中植入人耳听不到的频段的水印。由于施行令第23条第3款也认可通过软件进行确认，因此同时采用不可听水印和可听引导语是降低法律风险的履行方式。

Suno AI创作的歌曲提出了一个稍微不同的问题。AI作曲的歌曲是否属于“难以与真实情况区分的虚拟声音”，这存在解释的空间。如果是模仿真实歌手声音的歌曲，则适用第3款。如果是用虚拟声音创作的歌曲，虽然适用第2款的一般标识义务，但其是否符合第3款“难以与真实情况区分”的要件尚不明确。判断可能会根据模仿现存某人的声音或演奏风格的程度而有所不同。

中国的《深度合成管理规定》列举了六种技术类型：文本的生成和编辑、语音的生成和编辑、音乐和场景声等非语音声音的生成和编辑、图像和视频人脸等生物特征的生成和编辑、图像和视频非生物特征的生成和编辑、数字人与虚拟场景的生成和编辑。其中，语音属于单独的项目，而音乐等非语音声音属于另一个单独的项目。这与韩国法律第31条第3款将“声音、图像或视频”这三种媒介捆绑在单一条款中形成对比。中国按媒介进行了细分，而韩国则将其合而为一。

6. API提供者与使用API的服务：标识责任归谁

提供OpenAI API、Anthropic API、Google的Gemini API的公司是开发经营者。即第2条第7项甲目：‘开发并提供人工智能者。’而调用该API来制作自身服务的公司是使用经营者。即第2条第7项乙目：‘利用甲目经营者提供的人工智能来提供人工智能产品或人工智能服务者。’

第31条规定的透明度义务主体是“人工智能经营者”。无论是开发经营者还是使用经营者，都属于人工智能经营者。因此，两者都承担透明度义务。

作为开发经营者的API提供者，其义务有两个维度。在自己直接向使用者提供服务的情况下（如ChatGPT网页界面、Claude应用），直接履行事先告知和生成物标识义务。而在仅提供API的情况下，为使用经营者履行标识义务提供技术手段，更接近开发经营者的职责。例如，OpenAI在API响应中默认插入C2PA元数据，或者在生成的成果物中包含来源识别代码。

使用经营者的义务产生于其自身服务的背景下。如果是调用OpenAI API来运营客服聊天机器人的金融科技公司，则必须在聊天机器人屏幕上添加事先告知内容，并标明聊天机器人给出的回答是由AI生成的。即使API提供者已经进行了元数据级别的标识，使用经营者也必须另外履行在自身服务的使用者可见的层面上进行告知和标识的义务。

该法第34条第3款规定了防止双重负担的条款。虽然该内容仅限于高影响人工智能经营者确保安全性和信赖性的措施，但其主旨是，如果根据其他法令履行了相当于第34条第1款各项的措施，则视为已履行。然而，第31条规定的透明度义务中，并没有明确规定这种防止双重负担的条款。在API提供者与使用API的服务之间，标识义务的具体分配标准是今后需要通过指南或有权解释来填补的空白。

欧盟《人工智能法案》第50条对此问题的处理更为明确。生成合成声音、图像、视频、文本的AI系统的提供者（provider），包括通用AI系统在内，必须在其生成的成果物上进行机器可读格式的标识（第50条第2款），并且使用生成深度伪

造系统的部署者（deployer）也必须公开属于深度伪造的事实（第50条第4款）。这是根据价值链的各个阶段分配了定制化的义务。而韩国法律将义务强加于“人工智能经营者”这一单一主体，因此API提供者与API使用者之间的责任界限相对模糊。

7. 平台经营者的双重地位

Naver、Kakao、Google等平台经营者可能既是开发经营者，同时又是使用经营者。

如果Naver基于自主开发的HyperCLOVA X提供搜索AI服务，那么对于HyperCLOVA X而言，它就是开发经营者。如果同一个Naver引入OpenAI的模型来运营独立的翻译服务，那么对于该服务而言，它就是使用经营者。这两种地位是按服务进行划分的，并非指一个法人只能拥有一种地位。

这一双重地位在实务中发生冲突的地方，在于透明度义务的范围。

作为开发经营者的Naver，承担着对HyperCLOVA X的根本性职责：模型自身的安全性、训练数据的管理、以及内置生成物标识技术。而作为使用经营者的Naver，则对其引入的外部AI，履行自身服务背景下的透明度告知义务。

还有更复杂的案例。平台在开发自身AI模型的同时，将该模型以API形式提供给外部开发者，同时外部开发者利用该API制作服务并上架到平台应用市场（Marketplace）。如果Kakao开发了KoGPT并开放了API，而外部开发公司利用该API制作了聊天机器人并在KakaoTalk频道提供服务，那么Kakao对于KoGPT而言是开发经营者，外部开发公司是使用经营者。作为KakaoTalk频道这一分发平台的运营商，Kakao是否承担追加义务，这是法律未明确规定的问题。这与欧盟《人工智能法案》将分销商（distributor）和产品制造商（product manufacturer）规定为不同的受规范者形成对比。

判断是否为高影响人工智能也因服务而异。Naver的搜索人工智能服务很可能不属于高影响人工智能。因为它不直接属于第2条第4项列举的十一个领域，如能源供应、医疗保健体系或犯罪调查生物识别等领域。但是，如果同一个Naver开发了使用人工智能的招聘评估工具，就会触及第2条第4项G目“对招聘、贷款审查等个人权利和义务关系产生重大影响的判断或评估”。作为高影响人工智能经营者，将增加第34条确保安全性和可靠性的义务。制定风险管理方案、实施说明方案、运营使用者保护方案、人员管理与监督、文件的制作与保存。一项服务将承担这五项义务。

平台经营者有可以寻求咨询的地方。《人工智能基本法》支持服务台由科学技术信息通信部和韩国人工智能·软件产业协会（KOSA）共同运营。截至施行第10周，共接到552件咨询。关于透明度标识义务的咨询最多，占51%。可以进行商业秘密保护和匿名咨询。

8. 中国的事先登记义务：不同的监管哲学

韩国法律中没有在推出生成式人工智能服务前必须向政府登记的规定。中国则有。

中国《生成式人工智能服务管理暂行办法》于2023年8月15日起施行。该规定指出，在中华人民共和国境内向公众提供具有“舆论属性或者社会动员能力”的生成式人工智能服务，必须申请安全评估并完成算法备案。算法备案是根据《互联网信息服务算法推荐管理规定》在提供服务之日起10个工作日内进行申报的程序，而生成式人工智能模型的上线备案在实务中则要求在提供服务前完成。

登记现状说明了其规模。据悉，截至2024年8月，已有190多个大型语言模型完成了登记。根据中国国家互联网信息办公室的公告，截至2025年3月底，共有346个生成式人工智能服务完成备案，到2025年11月初增至611个。

用一句话总结韩国和中国的区别如下：在韩国，无需向政府进行事先登记或获得许可即可开始服务。在中国，必须履行登记义务后才能开始服务。

这种差异是监管哲学的差异。韩国《人工智能基本法》倾向于事后自主监管。由经营者自行评估是否属于高影响人工智能，并在必要时向科学技术信息通信部部长申请确认（第33条第1款）。这不是事先许可。中国则将事先登记与内容控制相结合。由国家互联网信息办公室（CAC）事先掌握服务的存在，并进行事后执法。

执法的力度也不同。韩国《人工智能基本法》的行政罚款上限为3000万韩元。第43条第1款规定，对违反事先告知义务、未指定韩国境内代理人、不履行中止命令或纠正命令的行为，处以3000万韩元以下的行政罚款。中国《生成式人工智能服务管理暂行办法》援引了现行《网络安全法》、《数据安全法》、《个人信息保护法》的制裁规定。在违反个人信息相关规定情节严重的情况下，根据《个人信息保护法》，最高可处以5000万元人民币（约合100亿韩元）的罚款，甚至可以责令停业整顿或停止提供服务。2025年，中国国家互联网信息办公室开展了“清朗”专项整治行动，对未完成备案的生成式人工智能应用程序和微信小程序进行了整顿。

在中国的监管中，韩国经营者还有一个需要注意的地方。2025年9月1日施行的《人工智能生成合成内容标识办法》规定，所有人工智能生成内容必须同时添加显式标识（肉眼可见的标签）和隐式标识（元数据或数字水印）。强制性国家标准GB 45438-2025规定了技术规范。这与韩国法律将“人类可识别的方法”或“机器可读取的方法”作为选项提供形成了对比。如果是有意或计划向中国市场提供生成式人工智能服务的韩国经营者，从一开始就在设计中纳入双重标识体系是合理的做法。

在API提供方面，中国的规定也很明确。《生成式人工智能服务管理暂行办法》在提供者的定义中写道：“包括通过可编程接口（API）等方式提供服务的情况”。API提供者作为服务提供者也承担备案义务和内容安全义务。韩国法律虽然将API提供者分类为开发经营者，但并没有对提供API本身施加单独的登记义务。

在韩国法律的事后自主监管和中国法律的事先登记控制之间，是欧盟的立场。欧盟《人工智能法案》（EU AI Act）对高风险人工智能实施了事先的合格评定（conformity assessment），但对于通用人工智能模型，则从被称为实践守则（Code of Practice）的软性规范开始。这三个国家（地区）的监管格局正以不同的速度发展。

9. 经营者现在需要确认的事项

将生成式人工智能用于业务的经营者，需要对自己的服务进行四方面的检查。

自己是开发经营者还是使用经营者。如果直接开发或微调人工智能模型并提供，则是开发经营者。如果调用外部人工智能来创建自己的服务，则是使用经营者。两者都可能适用。

自己服务的生成物是文本、图像、视频还是语音。文本生成服务适用第31条第1款和第2款。对于生成难以与真实情况区分的图像、视频和语音的服务，将增加第3款的更高标准。

自己的服务是否属于高影响人工智能领域。如果是触及第2条第4项列举的十一个领域的服务，则需要承担第34条的确保安全性和可靠性的义务。

如果是海外经营者，是否符合指定韩国境内代理人的标准。施行令第29条第1款规定的四项标准：上一年度销售额在1万亿韩元以上、人工智能服务部门上一年度销售额在100亿韩元以上、最近3个月日均韩国境内使用者人数在100万人以上、因不履行中止命令或纠正命令而被处以行政罚款的记录。只要满足其中一项，就必须指定韩国境内代理人，并向科学技术信息通信部部长申报。

回到位于江南办公室的那家房地产初创企业。该公司是使用经营者。它生成文本。房地产房源说明文字不是“难以与真实情况区分的虚拟语音、图像或视频”。该公司的义务是第31条第1款的事先告知和第2款的生成物标识。只需在服务画面上告知基于人工智能运营的事实，并在生成的文字上标注由人工智能生成即可。

即使OpenAI在API层面上履行了标识义务，该公司也必须另外进行使用者在其服务中可见的标识。法律质问的不是“有没有人做过”，而是“使用者能否知晓”。

第30章. 公共部门供应经营者

2026年春，首尔九老数码园区的一家AI初创企业代表打开了韩国调拨厅的Nara Marketplace（国家综合电子采购系统）。政府部门和地方自治团体发布的AI相关服务公告有15件。教育部发包了构建AI学生评估辅助系统，国民健康保险公团预告了引入健康体检影像判读AI。该代表的公司开发基于自然语言处理的信访自动分类系统。在一家地方自治团体有试点引入的业绩。然而，公告中出现了一个陌生的表述：“根据《人工智能基本法》第16条第3款，优先考虑科学技术信息通信部部长确认的人工智能产品或人工智能服务。”未获得该“确认”的产品也可以参与投标，但将被排除在优先考虑对象之外。如果与获得确认的竞争对手产品并列，就会处于劣势。这位代表翻开了施行令。

1. 优先考虑：法律创造的公共采购通道

《人工智能基本法》第16条第3款作了如下规定：“国家机关等在购买业务执行所需的产品或服务，或发包服务时，应优先考虑总统令规定的人工智能产品或人工智能服务。”附带了一个但书：“因业务特性不适合运用人工智能技术的情况除外。”

该条款作为2026年1月20日公布的修订法（法律第21311号）被添加，并自2026年7月21日起施行。这是原法中没有的条款。国会为了促进AI产业而将其加入。在立法过程中，它以“促进公共部门引入AI”的名义被讨论。

“应优先考虑”是一项义务条款，而非裁量。当国家机关、地方自治团体、根据《关于公共机关运营的法律》第4条规定的公共机关购买物品或提供服务时，如果AI产品或服务适合该业务，则必须首先考虑AI。但这并非强制购买义务。在考虑后若判断为不适合，可以选择其他方式。

施行令第15条第4款是关键。“总统令规定的人工智能产品或人工智能服务”是指“科学技术信息通信部部长确认的产品或服务”。必须获得科学技术信息通

信部部长的确认，才能成为优先考虑对象。确认的标准和程序由科学技术信息通信部部长与行政安全部部长协商后予以公告（施行令第15条第5款）。

对于试图向公共采购市场供应AI的经营者而言，该确认实际上是准入资格。即使法律上并非强制购买，但在获得确认的产品与未获得确认的产品竞争时，公共机关的经办人会优先选择已确认的产品。因为法律明文规定“应优先考虑”。虽然可以在没有确认的情况下参与投标，但由于被排除在优先考虑对象之外，可能会处于不利地位。

2. 确认与检认证：两道关卡

公共部门AI供应经营者面前有两道关卡。

第一道关卡是科学技术信息通信部部长的确认。以施行令第15条第4款为依据。获得此确认后，将成为所有公共机关采购中的优先考虑对象。确认的标准和程序由科学技术信息通信部部长与行政安全部部长协商后通过公告规定（施行令第15条第5款）。在政府发布阶段，曾提出由相关协会负责受理申请和发放确认书，由韩国信息通信技术协会（TTA）负责技术审查的方案，但确定的执行机构和详细程序须通过公告予以确认。

韩国人工智能振兴协会是根据《人工智能基本法》第26条设立的。它是从事AI相关研究和业务的人员经科学技术信息通信部部长认可而设立的法人。协会的法定业务包括促进AI技术和产品的使用、现状调查、专业人才培养、海外进军支持。在公共采购确认过程中，如果协会承担判断技术应用与否的角色，该协会将成为经营者们进入采购市场的守门人。

第二道关卡是检认证。以法律第30条为依据。科学技术信息通信部部长为了确保AI的安全性与可靠性，支持民间自主推进的验证和认证活动。支持内容包括普及指南、研究支持、设备和系统构建、专业人才培养。对于中小企业，可以支持检认证费用的全部或部分（施行令第22条第3款）。

该检认证在公共部门发挥效力的依据是第30条第4款。“国家机关等在试图使用高影响人工智能时，应优先考虑基于获得检认证等的人工智能的产品或服务。”若要向公共机关供应高影响人工智能领域的产品，获得检认证的产品优先。

整理这两道关卡的关系。若要以一般AI产品进入公共采购领域，科学技术信息通信部部长的确认是核心（第16条第3款）。若要作为高影响人工智能产品向公共机关供应，在此基础上还需要检认证（第30条第4款）。经营者应首先判断自己的产品是否属于高影响人工智能，如果属于，必须获得检认证才能成为完全的优先考虑对象。

第30条第3款对提供高影响人工智能的经营者规定了“应努力事先获得检认证等”。这是努力义务，并非强制。然而，由于第30条第4款强制要求国家机关优先考虑获得检认证的产品，对于试图进入公共市场的经营者来说，努力义务实际上成为了竞争条件。法律的文本是柔和的，但市场的逻辑是强硬的。

3. 48个机关的CAIO：采购决策的诞生地

《人工智能基本法》施行后，必须了解公共部门的AI采购决策是在哪里做出的。2025年9月18日，48个中央行政机关（部、处、厅）全部指定了人工智能首席责任官（CAIO, Chief AI Officer）。部长级机关的CAIO由副部长级担任，副部长级机关的CAIO由室长级担任。同日，CAIO协议会成立，并每月召开例会。

CAIO统筹该机关的AI引入战略。判断在哪些业务中引入AI、使用何种水平的AI、分配多少预算。AI经营者若要进入政府采购市场，其产品最终必须符合该CAIO的优先顺序。

CAIO协议会与国家人工智能战略委员会联动运作。战略委员会由作为委员长的总统1人、当然委员18人（16个中央行政机关的长官、国家安保室负责AI业务的次长、总统秘书室辅助AI业务的首席秘书官）、委任民间委员最多41人组成，总人数在60人以内，且委任委员必须占全体的半数以上（法律第7条第2款）。3名副委员长并非单独编制，而是由总统从当然委员或委任委员中指派，其中1名为

常驻。战略委员会确定了《大韩民国人工智能行动计划（基本计划2026~2028）》，这是一份包含99个执行课题和326项政策建议的199页文件。CAIO协议会协调该行动计划在各机关的履行。

对AI经营者而言，实务上重要的结构如下。战略委员会确定国家层面的AI政策方向，CAIO协议会将其传达给48个机关，各机关的CAIO确定本部门的AI引入优先顺序。调拨厅在Nara Marketplace发布公告后，经营者进行投标。这一流程的最顶端是战略委员会，而实务决策由CAIO做出。

试图承揽公共部门AI项目的企业常常存在一个误区。那就是只盯着调拨厅。调拨厅是管理采购程序的机关，而非决定买什么的机关。买什么是由各部门的CAIO和实务部门决定的。经营者必须在采购公告发布之前，掌握相关部门正在计划哪些AI课题，以及CAIO协议会正在讨论什么方向。

2026年6月30日，人工智能政策中心正式启用。韩国智能信息社会振兴院（NIA）根据法律第11条被指定（迄今为止被指定的仅有NIA），由3个组和1个团，共5个小组、59人组成。支持基本计划制定、支持政策开发、AI扩散影响调查与分析、法律制度研究是其法定业务。该中心扮演着公共部门AI引入的实务枢纽角色。经营者可以通过该中心发行的引入案例集和技术现状报告，解读公共部门的需求。

4. 公务员免责：开启大门的钥匙

难以向公共部门销售AI的原因之一是公务员的恐惧。如果引入了AI系统却发生故障引发信访，批准引入的经办人要承担责任。可能会在监察院的审计中受到指责，甚至被追究赔偿责任。安全的选择是不引入AI。如果维持现有方式，就没有人会追究。

《人工智能基本法》第16条第4款改变了这种结构。“即使因购买或使用根据第3款规定的人工智能产品或人工智能服务给国家机关等造成损失，负责该购买或使用业务的人员也不对该机关的损失承担赔偿责任。但是，因该业务经办人的故意或重大过失造成损失的情况除外。”

让我们剖析一下该条款的结构。保护对象是“负责购买或使用业务的人员”。即批准引入和运营AI的公务员。免责范围是“对该机关损失的赔偿责任”。这意味着当AI导致机关发生金钱损失时，不需要经办公务员来赔偿这笔钱。但书条款是“故意或重大过失”。如果明知危险却引入AI，或者极其疏忽大意，则被排除在免责之外。

此免责存在前提条件。必须是“根据第3款规定”的产品或服务。只有在按照优先考虑程序购买或使用科学技术信息通信部部长确认的AI产品或服务的情况下，才适用免责。如果随意购买未经确认的AI产品而发生问题，将很难获得此免责的保护。

从经营者的立场来看，为什么这个免责条款很重要呢。如果公务员对引入AI的恐惧减少，采购决策就会加快。这等于是法律对“引入AI后如果发生问题怎么办”的问题给出了答案。如果正常引入了科学技术信息通信部部长确认的产品，且不存在故意或重大过失，经办人就不承担责任。

这种结构也给经营者提供了间接保护。如果公务员不怕购买AI，市场就会打开。市场一旦打开，就会产生B2G销售额。免责虽然是为公务员制定的条款，但其效果直接关系到经营者的销售额。

但是，此免责免除的是公务员对机关的追偿责任，而不是免除国家对国民承担的国家赔偿责任。如果由于AI系统的故障给国民造成损失，国家必须根据《国家赔偿法》进行赔偿。也有向开发AI的经营者行使追偿权的余地。经营者不应因为有免责条款就掉以轻心，而是必须确保产品本身的安全性。

5.（子）目：公共机关利用AI进行决策时

向公共部门供应AI的经营者必须了解一个条款。即第2条第4项（子）目。这是高影响人工智能的11个领域中直接适用于公共部门的项目。

（子）目的原文如下：“国家、地方自治团体、根据《关于公共机关运营的法律》第4条规定的公共机关等，在提供公共服务所需的资格确认及决定、或费用征收等对国民产生影响的决策。”

我们来剖析这句话。主语是国家、地方自治团体、公共机关。行为是决策。决策的内容是资格确认、资格决定、费用征收。因为带有“等”字，所以这三项是示例。

代入具体案例。如果在公务员考试阅卷中使用AI，就符合（子）目。因为这是资格确认兼决定。如果由AI判定福利待遇领取资格，就是（子）目。如果AI介入税金计算，就属于费用征收。在健康保险费计算、水费定价、居民注册相关行政决策中使用AI的情况也被涵盖在内。

如果向公共机关供应符合（子）目的AI系统，该系统就会成为高影响人工智能。高影响人工智能经营者的责任也会随之而来。根据第34条，需要制定风险管理方案、制定解释方案、制定使用者保护方案、确保人工管理监督体系、起草及保管文件。

经营者需要注意一点。（子）目的主语是公共机关，而不是经营者。由公共机关进行决策。然而，制造用于该决策的AI系统的经营者，即开发经营者要承担高影响人工智能的义务。公共机关根据使用方式的不同，可能成为使用者，也可能成为经营者。第2条第7项将国家机关等包含在人工智能经营者的定义中。当公共机关将从外部购买的AI用于自身行政时，该机关就成为使用经营者。

结果是，向公共机关供应AI的开发经营者，以及引入AI用于行政的公共机关（使用经营者），双方都成为高影响人工智能义务的规范对象。但是，如果开发经营者履行了风险管理方案、解释方案、使用者保护方案（第34条第1款第1项至第3项），只要作为使用经营者的公共机关没有进行重大的功能变更，就在该范围内视为履行了义务（施行令第27条第3款）。人员的管理和监督、起草和保管文件的义务，由公共机关单独履行。这是为了减少双重负担的机制。

还有一点。如果公共机关已经根据总统令的规定，依照其他法令采取了同等措施，则视为履行了基本法的义务（第34条第3款，施行令第27条第5款及附表1）。例如，如果韩国国税厅在税务行政中使用AI，并且根据《国税基本法》或《电子政府法》具备了合法程序，则可能无需单独履行基本法的风险管理义务。

6. 影响评估这一武器

在公共采购市场中，影响评估既是义务也是武器。

《人工智能基本法》第35条允许人工智能经营者自主实施高影响人工智能影响评估。这不是强制性的。但是，履行该评估的经营者在公共采购中占据有利地位。这就是在立法过程中被称为“软监管”的结构。虽然法律没有直接强制，但通过与公共采购优先考虑挂钩，使其实际上发挥了市场准入条件的作用。

影响评估包含的内容由施行令第28条规定。包括七项：识别受影响的个人和群体、识别基本权类型、社会及经济影响、使用形态、评估指标和计算方式、风险的预防、缓解和损失恢复，以及改进履行计划。实务推进方面，《影响评估指南》分为事前准备、正式评估（编写风险情景和分析基本权影响）、事后措施三个阶段进行指导。经营者须将此评估形成文件并予以保管。

对于试图向公共部门供应AI的经营者而言，完成影响评估就如同获得了一份“我们的产品已经考量了对基本权的影响，并具备了管理风险的体系”的证明书。公共机关经办人在接受审计时，可以提出该文件作为依据。由于能够通过文件展示引入AI的合理理由，影响评估也是公共机关的防御手段。

科学技术信息通信部于2026年1月22日发布的《AI影响评估指南》提出了该评估的实务标准。经营者按照该指南进行评估即可。也可以在AI基本法支持服务台（由韩国软件产业协会运营）进行匿名咨询。以施行第10周为准，已受理了552件咨询。

7. 国防与安保：法律之外的市场

《人工智能基本法》第4条第2款非常明确：“本法不适用于仅出于国防或国家安保目的而开发、使用且由总统令规定的人工智能。”

“仅出于”三个字是核心。只有在军用专有、安保专有的时候才排除适用。军民两用（dual-use）不被排除。即使是向军队供应的AI，如果是民间也能使用的通用技术，则适用基本法。

施行令第2条限定了适用排除范围。有三个指定主体。

国防部部长指定的业务。包括国防信息通信网和国防信息系统的构建与运营、武器体系和战力支持体系的开发与运用。

这是国家情报院院长指定的业务。列举了经济安保预警系统运营、安保课题研究、网络安全预防与应对体系构建、国家尖端战略技术泄露调查、反恐活动、国防工业技术保护、防止产业技术泄露、电子政务安全措施、综合防卫情报业务、反间谍业务、韩国《国家情报院法》上的业务。

这是警察厅厅长指定的业务。包括尖端战略技术泄露调查、反恐活动、国防工业技术泄露调查、核心技术泄露调查、反间谍业务、情报犯罪调查。

该目录为限制性列举。未包含在此处的国防或安保相关人工智能受《人工智能基本法》适用。军队餐饮管理人工智能、军事设施冷暖气优化人工智能、军人健康管理人工智能等不属于战斗或安保目的，因此不属于排除适用的对象。

对于试图进入国防人工智能采购市场的经营者而言，这一区分非常重要。如果开发属于排除适用对象的人工智能，则不受高影响判断、影响评估、透明度义务等《人工智能基本法》义务的约束。取而代之的是，适用韩国《防卫事业法》、《军事机密保护法》等国防相关法律。排除适用并不意味着免除监管，只是适用的法律发生变化而已。

EU AI Act也具有类似的结构。第2条第3款排除了“仅（exclusively）出于军事、国防或国家安全目的推出或使用的人工智能系统”的适用。双方都使用了“专用（exclusively）”这一限定词。如果说有差异，那就是欧盟的法律条文本身是自动执行的（self-executing），而韩国则将具体范围授权给总统令。韩国的施行令第2条毫无遗漏地列举了业务。未列举的业务很难主张排除适用。

8. 欧盟与中国的公共人工智能采购

将韩国的公共部门人工智能优先考虑制度与外国进行比较。

EU AI Act并未单独规定促进公共部门引入人工智能。欧盟的法律是监管法。产业振兴通过单独的政策来推进。欧盟通过数字欧洲计划（Digital Europe Programme）支持公共部门引入人工智能，但并未在人工智能法本身加入“公共机构应优先购买人工智能”的条款。当欧盟的公共机构引入高风险人工智能时，必须强制实施基本权利影响评估。这不是促进条款，而是监管条款。

韩国将促进与监管混合在同一部法律中。第16条第3款是促进。第34条是监管。从经营者的立场来看，韩国比欧盟更容易进入公共采购市场。因为法律鼓励公共机构购买人工智能。欧盟的符合性评估（conformity assessment）是强制性的监管要求，只要是高风险人工智能，无论是否采购都必须通过。韩国的确认与验证及认证是一旦获得便随之带来优先考虑这一优惠的任意性优待机制。这两种制度在门槛的性质上本身就不同。

中国没有关于公共部门引入人工智能的综合性条款。各部门单独推进人工智能试点项目，并以CAC（国家互联网信息办公室）为中心，由9个以上部门按领域进行监管。《生成式人工智能服务管理暂行办法》第17条要求具有舆论属性或社会动员能力的生成式人工智能服务进行安全评估和算法备案。面向大众的服务大部分都属于此类，因此，无论是公共部门还是民间，如果没有备案，都很难提供这种服务。截至2025年11月，已有611个生成式人工智能服务完成备案。要想进入中国的公共部门人工智能市场，这一备案实际上是先决条件。这与韩国的自主

确认程序性质不同。

从这三个国家的公共部门人工智能采购方法来看，存在明显的差异。韩国是“法律推动政府购买人工智能。”欧盟是“如果政府要购买人工智能，必须满足法律规定的安全要求。”中国是“如果是能够影响舆论的人工智能服务，无论是政府还是民间，都必须先备案。”

9. 采购程序的实务

接下来了解公共部门人工智能采购的实际流程。

韩国调达厅的国家综合电子采购系统（G2B）是公共采购的中心平台。人工智能相关的采购主要分为三种途径。

第一是普通招标。发包机构确定规格，在国家综合电子采购系统上发布公告，经过资格审查或技术评估后确定中标者。人工智能项目也遵循这一途径。但是，在《人工智能基本法》施行后，科学技术信息通信部部长确认的产品或服务可能会被赋予加分或优先考虑。

第二是多重供应者合同（MAS）。即韩国调达厅事先与满足一定质量标准的多数供应者签订合同，需求机构从中挑选购买的方式。人工智能解决方案可以登记在软件、信息系统类别中。

第三是创新产品试点购买项目。即政府对市场上尚未出现或应用了新技术的试用产品进行购买，并在公共机构进行测试的制度。这是人工智能初创企业迈向公共市场的第一步可以使用的途径。

科学技术信息通信部每年运营“人工智能代金券项目”。该项目旨在向中小企业发放代金券（使用券），以便其引入人工智能解决方案，在此过程中会选定人工智能供应企业。虽然这不是公共机构直接引入，但发挥了验证人工智能供应企业业绩与能力的功能。在代金券项目中积累的业绩将成为公共采购招标的参考资料。

10. B2G人工智能市场的地形

韩国公共部门人工智能市场在《人工智能基本法》施行后具备了结构性扩张的条件。

存在法律上的需求创造机制。第16条第3款的优先考虑义务鼓励公共机构购买人工智能。48个CAIO（首席人工智能官）负责推进各机构引入人工智能。公务员免责机制降低了做出引入决定的心理障碍。

预算途径已打开。推进数字平台政府的预算被分配给人工智能基础设施。行政安全部主管公共机构的数字化转型，科学技术信息通信部负责人工智能技术的普及。施行令第15条第2款规定这两个部门应通过协商制定引入人工智能支援的具体方案。

已经有正在运行的人工智能引入案例。国税厅的税务调查对象选定人工智能、关税厅的进口物品检索人工智能、警察厅的闭路电视分析人工智能、政府24的信访向导人工智能聊天机器人、国民健康保险公团的健康检查影像判读辅助人工智能都已经在现场运行。这些都是在法律施行前引入的。法律施行后，需要重新检查这些系统是否满足《人工智能基本法》的义务，而新系统的引入将朝着具备确认与验证及认证的方向进行。

但是，即使市场开放，对所有人工智能经营者而言机会也并不均等。在科学技术信息通信部部长确认程序、验证及认证费用、执行影响评估的能力方面，大企业和初创企业之间存在差距。该法第30条第2款规定要向中小企业支援验证及认证费用，施行令第22条第3款写明可以在预算范围内支援全部或部分费用。这一支援实际达到何种水平，将决定中小人工智能经营者进入公共市场的步伐。

将公共部门人工智能项目用一句话概括就是这样：获得科学技术信息通信部部长确认，判断是否属于高影响，如果属于，则具备验证及认证以及影响评估，然后向国家综合电子采购系统提交履历。阅读该履历的人是48个机构的CAIO。

还有一个遗留的问题。《人工智能基本法》的行政罚款上限为3000万韩元，且仅在违反事先告知、未指定韩国境内代理人、未履行中止或纠正命令这三种情况下才会处以行政罚款（第43条第1款）。违反第34条的责任并不会直接导致行政罚款，而是经过事实调查和纠正命令，在不履行该命令时才会导致行政罚款。EU AI Act的罚款上限为3500万欧元（约500亿韩元）或全球销售额的7%。然而，在公共采购市场中，实际成本并非行政罚款。如果确认被取消，或者在采购过程中失去信任，可能会在其他机构的采购中被排挤。公共采购市场是一个退出市场比金钱制裁更可怕的地方。

附录

附录 A. 经营者自检流程图

本流程图以韩国《人工智能基本法》（根据法律第20676号制定，根据法律第21311号修改，修改条款于2026年7月21日施行）及其《施行令》（根据总统令第36053号制定，反映了2026年7月20日公布并于次日施行的修改内容）为准编制。经营者在判断自身服务的法律地位和义务时，依序参考即可。

判定共分为六个步骤。

判定 1. 是否属于人工智能系统

依据条款：本法第2条第2项

提问：贵公司的系统是否同时具备以下三项要素？

(a) 具有自主性和适应性。根据输入数据自行调整判定标准。

(b) 通过推理得出输出结果。执行预测、推荐、决定中的一项或多项。

(c) 对现实或虚拟环境产生影响。

是：进入判定 2。

否：不属于《人工智能基本法》的适用对象。在此结束。

参考：基于规则的 RPA、if-then 宏以及仅靠固定公式运行的软件因不具备自主性和适应性，不属于人工智能系统。

相反，如果机器学习模型是服务的核心引擎，则很有可能属于人工智能系统。

判定 2. 是否属于人工智能经营者

依据条款：本法第2条第7项

提问 2-1：贵公司是否直接开发人工智能并向外部提供？

是：属于开发经营者。进入判定 3。

提问 2-2：贵公司是否利用其他经营者开发的人工智能向外部提供产品或服务？

是：属于使用经营者。进入判定 3。

提问 2-3：贵公司是否仅接受人工智能产品或服务，而不再以此为基础向外部提供产品或服务？

是：属于使用者（本法第2条第8项）。无需承担法律规定的经营者义务。

在此结束。

参考：可同时具备两种身份。如果自行开发人工智能，并利用该人工智能运营贷款审批服务，则既是开发经营者，又是使用经营者。

法人、团体、个人、国家机关均可适用，且不论是否以营利为目的。

参考：对外部模型进行微调并作为新服务推出的情况，因属于实质性的开发行为，可被归类为开发经营者。

判定 3. 是否属于生成式人工智能

依据条款：本法第2条第5项

提问：贵公司的人工智能系统是否通过模仿输入数据的结构和特征，生成文本、声音、图像、视频等新的输出结果？

是：属于生成式人工智能。将增加以下义务：

义务 (1) 生成物标识：必须以使用者可识别或机器可读取的方式，标识该输出结果为AI生成。（本法第31条第2款、施行令第23条第2款）

义务 (2) 事先告知：必须预先向使用者告知该产品或服务是基于人工智能运行的事实。（本法第31条第1款、施行令第23条第1款）

义务 (3) 深度伪造告知与标识：如果提供难以与真实情况区分的虚拟声音、图像、视频等，必须以使用者能够清晰识别的方式进行告知或标识。这是与是否属于生成式人工智能无关的独立义务，对于艺术性、创意性的表现物，可以采用不妨碍展示或欣赏的方式进行。（本法第31条第3款）

进入判定 4。

否：不属于生成式人工智能。

虽然不适用生成式人工智能的相关标识义务，但如果属于高影响人工智能或高性能人工智能，则会产生其他义务。进入判定 4。

参考：典型代表如 ChatGPT、Midjourney、Suno AI 等生成文本、图像、音乐、视频的系统。仅进行分类预测而不生成新内容的系统不属于此列。

判定 4. 是否属于高影响人工智能

依据条款：本法第2条第4项、施行令第25条（高影响人工智能的确认程序等）

高影响人工智能的判定分为两个步骤。

步骤 4-1. 领域确认

贵公司的人工智能系统是否用于以下 11 个领域中的一个或多个？

(a) 能源供应

(b) 饮用水生产流程

(c) 医疗保健提供与使用体系的构建和运行

(d) 医疗器械及数字医疗器械的开发与使用

(e) 核材料及核设施的安全管理与运行

(f) 用于侦查犯罪或逮捕的生物识别信息（人脸、指纹、虹膜、掌静脉等）的分析与应用

(g) 招聘、贷款审批等对个人的权利和义务产生重大影响的判定或评估

(h) 交通工具、交通设施、交通体系的主要操作与运行

(i) 涉及公共服务资格确认、决定、费用征收等对国民产生影响的国家机关决策

(j) 幼儿教育、初等教育、中等教育中的 student 评估

(k) 其他由施行令规定的领域

否：不属于高影响人工智能. 进入判定 5。

是：进入步骤 4-2。

步骤 4-2. 重大性判定

相关人工智能系统是否有可能对人的生命、人身安全或基本权利产生重大影响，或者带来风险隐患？

否：不属于高影响人工智能. 进入判定 5。

是：属于高影响人工智能. 适用以下义务。

高影响人工智能经营者的主要义务：

义务 (1) 制定和运行风险管理方案：必须建立风险识别、分析、评估、处理和改进的体系。（本法第34条第1款第1项）

义务 (2) 制定和实施解释方案：应就最终结果、主要标准、训练数据概要拟定解释方案，并将主要内容公示在办公场所或官方网站等处。（本法第34条第1款第2项、施行令第27条第1款）

义务 (3) 制定和运行使用者保护方案（本法第34条第1款第3项）

义务 (4) 人工管理与监督：必须设立允许人工干预人工智能决策的机制。（本法第34条第1款第4项）

义务 (5) 文档化：履行依据文件须保存 5 年。（本法第34条第1款第5项、施行令第27条第2款）

义务 (6) 影响评估：应努力自主评估对基本权利产生的影响。（本法第35条）

进入判定 5。

参考：可向科学技术信息通信部长官申请确认是否属于高影响人工智能. 确认期限为 30 天，可延长一次，延长天数不超过 30 天。（本法第33条、施行令第25条第3款）

判定 5. 是否属于高性能人工智能

依据条款：本法第32条、施行令第24条

提问：贵公司的人工智能系统是否同时满足以下三项条件？

(a) 用于训练的累计计算量在 10 的 26 次方浮点运算（FLOPs）以上。

(b) 采用并运行当前属于最前沿的人工智能技术。

(c) 可能会对人的生命、人身安全及基本权利产生广泛而重大的影响。

是：属于高性能人工智能. 适用以下义务：

义务 (1) 识别、评估及缓解贯穿整个生命周期的风险（本法第32条第1款第1项）

义务 (2) 建立旨在监测及应对安全事故的风险管理体系（本法第32条第1款第2项）

义务 (3) 向科学技术信息通信部长官提交履行结果（本法第32条第2款）

进入判定 6。

否：不属于高性能人工智能. 进入判定 6。

参考：必须同时满足三个要件. 如果仅计算量庞大而未能符合其他要件，则不属于高性能人工智能。

对比：欧盟《人工智能法案》（EU AI Act）的通用人工智能（GPAI）系统性风险标准为 10^{25} 次方 FLOPs. 由于韩国的标准高出 10 倍，因此韩国本土的大多数经营者不在此列。

判定 6. 是否承担指定韩国境内代理人的义务

依据条款：本法第36条、施行令第29条

本步骤仅适用于在韩国境内没有住所或营业场所的境外经营者。

若在韩国境内设有住所或营业场所，则跳过本步骤。

提问：是否符合以下四项中的任意一项？

(a) 上年度营业额在 1 万亿韩元以上。

(b) 人工智能服务领域的上年度营业额在 100 亿韩元以上。

(c) 最近 3 个月内，韩国境内的日均使用者人数在 100 万人以上。

(d) 曾有根据本法第43条第1款第3项被处以行政罚款的记录。

是：必须以书面形式指定韩国境内代理人，并向科学技术信息通信部长官申报。

韩国境内代理人必须是在韩国境内设有住所或营业场所的人。

韩国境内代理人代理的事项：

提交高性能人工智能履行结果（本法第32条第2款）

申请确认是否属于高影响人工智能（本法第33条第1款）

支持履行确保安全性与可靠性的措施（本法第34条第1款）

未指定时：处以 3000 万韩元以下的罚款（本法第43条）

否：无须承担指定韩国境内代理人的义务。

参考：根据施行令第29条第2款，境外经营者的营业额以按上年度平均汇率折算为韩元的金额为准。

自检结果汇总表

下表整理了根据判定结果所适用的义务组合。请确认相应的栏目。

普通人工智能经营者（在判定 2 中确认）

若不属于高影响、生成式、高性能中的任何一种，则无需承担本法规定的具体个别义务。事先告知义务（本法第31条第1款）仅在提供基于高影响或生成式人工智能的产品或服务时产生。

生成式人工智能经营者（在判定 3 中确认）

事先告知义务（本法第31条第1款）

生成物标识义务（本法第31条第2款）

深度伪造告知与标识义务：提供相关输出结果时（本法第31条第3款）

高影响人工智能经营者（在判定 4 中确认）

事先告知义务（本法第31条第1款）

制定和运行风险管理方案（本法第34条第1款第1项）

制定和实施解释方案（本法第34条第1款第2项）

制定和运行使用者保护方案（本法第34条第1款第3项）

人工管理与监督（本法第34条第1款第4项）

履行依据文件保存 5 年（本法第34条第1款第5项、施行令第27条第2款）

努力开展基本权利影响评估（本法第35条）

高性能人工智能经营者（在判定 5 中确认）

识别、评估和缓解风险（本法第32条第1款第1项）

建立风险管理体系（本法第32条第1款第2项）

向科学技术信息通信部提交履行结果（本法第32条第2款）

境外经营者韩国境内代理人（在判定 6 中确认）

书面指定并申报韩国境内代理人（本法第36条）

本流程图反映了截至2026年7月21日的现行法律（已反映修改后的法律第21311号）以及同日施行的修改后的施行令。您可以通过科学技术信息通信部《人工智能基本法》支持热线（www.sw.or.kr/AI_act_helpdesk）进行匿名咨询。

附录 B. 条文-施行令-指南对应表

将韩国《人工智能基本法》（作为第20676号法律制定并于2026年1月22日施行，作为第21311号法律修订，修订条文于2026年7月21日施行）的核心条文置于第1列，将具体化该条文的施行令（作为第36053号总统令制定，反映2026年7月20日公布并于次日施行的修订部分）条款置于第2列，将相关的科学技术信息通信部指南（五大指南及两件公告，2026年1月22日发布）置于第3列。

五大指南简称：

[透明性] 人工智能确保透明性指南

[安全性] 人工智能确保安全性指南

[高影响判断] 高影响人工智能判断指南

[经营者责任] 高影响人工智能经营者责任指南

[影响评估] 人工智能影响评估指南

两件公告简称：

[最前沿公告] 最前沿人工智能确保安全性公告

[责任公告] 高影响人工智能经营者责任公告

参考：本表的条文编号遵循以2026年7月21日为准在国家法令信息中心确认的现行法令编号。国会审查阶段的法案草案或部分解说资料中仍保留着其他编号（如第15条事先告知、第16条生成物标识等），因此旧编号与现行编号的对应关系将在备注1中整理。

条文（基本法） | 施行令（2026年7月21日施行现行） | 相关指南/公告

=====

=====

第2条第1项（人工智能）

“将学习、推理、感知、判断、语言的 | 施行令中无单独的具体化条款。 | 在[高影响判断]中作为AI适用性

理解等人类所具有的智力能力， | 法律本身为自我完备性定义。 | 判断标准予以援用。

以电子方法予以实现之物” ||

第2条第2项（人工智能系统）

具有各种水平的自主性和适应性， | 施行令中无单独的具体化条款。 | 在[高影响判断]中对于系统适用性

推理出预测、推荐、决定等 | 基于规则的自动化（RPA）不 | 判断时，将自主性、适应性要件

结果物的系统 | 包含在该定义中。 | 作为示例进行解说。

第2条第4项（高影响人工智能）

存在对人的生命、身体的安全性及 | 施行令第25条：高影响确认程序。 | [高影响判断] 全部。

基本权利造成重大影响或 | 提交确认请求书，30日（可延长30日） | 将两阶段判断（领域适用性 +

引发风险之虞的 | 内回复，重新确认请求于10日 | 重大性）以详细标准和

人工智能系统。列举11个领域 | 内提交。 | 案例进行解说。

(A至K目)。 | 施行令第26条：设立专业委员会。 |

第2条第5项（生成式人工智能）

模仿输入数据的结构和特性， | 施行令第23条第2款：生成物 | 在[透明性]中指南生成式人工智能

生成文字、声音、图像、视频 | 标识方法（人可识别方法 | 结果物标识方法。

等结果物的系统 | 或机器可读方法）。 |

第2条第7项（人工智能经营者）

A. 开发经营者：开发人工智能 | 施行令第27条第3款：开发经营者 | 在[经营者责任]中解说开发经营者与

并予以提供者 | 履行第34条第1款第1至3项措施的 | 使用经营者的义务界限。

B. 使用经营者：利用开发经营者 | 情况下，若使用经营者未进行重大功能 | 包含兼具两种地位的案例。

提供的AI来提供产品 | 变更，则视为在该范围内 |

或服务者 | 已履行。 |

| 施行令第27条第4款：使用经营者 |

| 可向开发经营者请求提供资料。 |

高性能人工智能

（基本法第2条中无定义。 | 施行令第24条：全部满足三项标准 | [安全性] 及 [最前沿公告]。

接受法第32条之委任， | 时适用。 | 详细指南运算量计算方式、

在施行令中予以规定) | 1) 累计运算量10的26次方FLOP以上 | 风险识别/评估/
缓解程序、

| 2) 适用最前沿技术 | 结果提交方法。

| 3) 存在对生命/身体/基本权利造成广泛且 |

| 重大影响之虞 |

第3条 (基本原则及国家等之责任)

第1款: 提高安全性、可靠性方向 | 施行令中无单独的具体化条款。 | 五大指南
全部

第2款: 受影响者的说明请求权 | 基本原则在法律层面上 | 均将该基本原则作为
背景规范

(在技术上、合理可能的 | 进行自我完备性宣告。 | 予以援用。

范围内提供明确且有意义的 | | 第2款说明请求权与

说明) | | [经营者责任]中的确立说明方案

第3款: 国家/地方自治团体安全使用环境 | | 义务相连结。

第4款: 应对社会变化之施政措施 | |

第4条 (适用范围)

第1款: 域外适用。 | 施行令第29条: 韩国境内代理人指定 | 无直接对应指南。

即便是境外行为，若对韩国境内市场或 | 标准（销售额1万亿韩元、AI销售额100亿韩元、 | 在[透明性]中提及海外经营者也

使用者造成影响，则予适用。 | 日均韩国境内使用者100万人、 | 承担透明性义务。

| 第43条行政罚款处分履历）。 |

第2款：国防/国家安全专用AI | 施行令第2条：除外适用AI范围。 |

除外适用。 | 国防部部长指定（国防信息通信网、 |

“仅以……为目的” = 双重用途是 | 武器系统），国家情报院院长指定 |

规制对象。 | （早期预警、反恐等11种类型）， |

| 警察厅厅长指定（反恐、尖端技术 |

| 侵害调查等6种类型）。 |

第31条第1款（事先告知）

在提供利用高影响或生成式AI的 | 施行令第23条第1款：告知方法4种。 | [透明性] 全部。

产品/服务时，将基于 | 1) 在产品上直接记载，在合同、 | 详细指南事先告知文句示例、

该AI予以运用之事实 | 使用说明书、使用条款中记载 | 告知位置、时点、例外

事先告知使用者。 | 2) 在使用者画面/终端机上显示 | 判断标准。

| 3) 在提供场所揭示 |

| 4) 部长认可的方法 |

| 施行令第23条第4款：例外。 |

| 明确的情况、内部业务专用、 |

| 部长公告例外。 |

第31条第2款（生成物标识）

标示生成式AI或利用该AI的 | 施行令第23条第2款：标识方法。 | 在[透明性]中按
类型

产品/服务的结果物是由生成式 | 1) 人可识别的方法 | （文本、图像、

AI生成之事实。 | 2) 机器可读的方法（此情况下 | 音频、视频）指南标识方法

| 提供1次以上提示文句/语音）。 | （水印、标签、元数据）。

第31条第3款（深度伪造标识）

在提供利用AI系统且与实际 | 施行令第23条第3款：以便于通过视觉、听觉 | 在[
透明性]中将深度伪造标识

难以区分的虚拟音响、 | 等易于确认， | 义务作为单独项目予以指南。

图像、视频等时，为使使用者 | 须考量使用者年龄、身体/社会条件 | 包含适用
艺术/创意表达物

能明确认知，须予以 | 进行标识。 | 例外的案例。

告知或标识。 ||

艺术/创意表达物可采用 ||

不阻碍展示/欣赏的 ||

方式进行标识。 ||

第32条（确保安全性义务，高性能AI）

第1款：累计运算量在施行令 | 施行令第24条第1款：全部满足三项标准 | [安全性] 及 [最前沿公告]。

标准以上的AI系统之确保安全性 | （运算量 10^{26} FLOP、最前沿技术、 | 指南
风险识别/评估/缓解履行

义务。 | 存在广泛/重大影响之虞）。 | 方式，风险管理体系构建

第1项：生命周期全般风险识别、 | 施行令第24条第2款：运算量计算 | 方法，
提交材料表格。

评估及缓解 | 方式由部长公告。 |

第2项：安全事故监控及 ||

风险管理体系构建 ||

第2款：将履行结果向科学技术信息通信部 ||

部长提交。 ||

第3款：部长公告履行方式。 ||

第33条（高影响人工智能的确认）

第1款：经营者的事先研讨义务。 | 施行令第25条：确认程序详细。 | [高影响判断] 全部。

必要时向部长请求确认。 | 提交确认请求书 + 产品概要书、 | 各11个领域的适用案例，

第2款：部长的确认义务， | 学习用数据概要、结果资料提交。 | 两阶段判断（领域 + 重大性）的

可向专业委员会咨询。 | 30日内回复（可延长30日）。 | 具体标准和示例。

第3款：制定/普及指南。 | 可于10日内请求重新确认。 |

第4款：程序事项委任于施行令。 | 施行令第26条：构成专业委员会 |

| （50人以上，相关公务员与AI技术/ |

| 伦理/法律及各领域专家， |

| 可设立咨询用5人小委员会）。 |

第34条（高影响AI经营者的责任）

第1款：确保安全性/可靠性措施。 | 施行令第27条第1款：将以下内容 | [经营者责任] 全部及

第1项：确立/运营风险管理方案 | 揭示于办公室/营业场所或主页上。 | [责任公告]。

第2项：确立/施行说明方案 | 1) 风险管理政策及组织体系 | 指南确立风险管理方案的5个阶段

（主要标准、学习数据概要） | 2) 说明方案主要内容 | （政策确立、风险识别、

第3项：使用者保护方案 | 3) 使用者保护方案 | 分析/评估、处理、改善），

第4项：人的管理/监督 | 4) 管理/监督者姓名及联系方式 | 说明方案构成要素、

第5项：文书拟定与保管 | （属于商业秘密的事项可除外） | 使用者保护程序、

第6项：委员会审议/表决事项 | 施行令第27条第2款：文书保管5年。 | 管理/监督者指定标准、

第2款：部长公告及建议。 | 施行令第27条第5款：依据其他法令 | 指南文书化项目和保管方法。

第3款：履行其他法令时 | 的视为履行（附表1）。 |

视为履行。 ||

第35条（影响评估）

第1款：提供使用高影响人工智能 | 施行令第28条：影响评估包含事项。 | [影响评估] 全文。

产品/服务时评估基本权利 | 1) 识别受影响的个人/群体 | 按步骤解析影响评估的

影响的“努力义务”。 | 2) 识别受影响的基本权利类型 | 3个阶段（事前准备、

第2款：国家机关等优先考虑已 | 3) 社会/经济影响的内容及范围 | 正式评估、事后措施）。

实施影响评估的产品/服务。 | 4) 使用行为模式 | 包含风险情景编写方法、

第3款：具体内容/方法 | 5) 定量/定性评估指标 | 基本权利影响分析方法、

委托给施行令。 | 6) 预防/缓解风险及恢复损失 | 自主评估运营案例。

| 7) 需要改善时的履行计划 |

| 可直接实施或委托第三方。 |

第36条（指定韩国境内代理人）

第1款：在韩国境内无住所/营业所 | 施行令第29条：指定标准。 | 无直接对应的指南。

的境外经营者中，使用者人数、 | 1) 上年度销售额在1万亿韩元以上 | 在[透明性]与[经营者责任]中

销售额等符合施行令标准的 | 2) 人工智能销售额在100亿韩元以上 | 部分提及了境外经营者的

经营者，须以书面形式指定 | 3) 直前3个月日均韩国境内使用者 | 义务。

韩国境内代理人并向长官申报。 | 在100万人以上 |

代理事项： | 4) 曾被处以第43条行政罚款的 |

第1项：提交第32条的履行结果 | 经营者 |

第2项：提出第33条高影响确认申请 | 境外销售额按上年度平均汇率折算为 |

第3项：支持履行第34条 | 韩元。 |

确保安全性/可信度措施 ||

第3款：代理人违规时，视为指定 ||

该代理人的经营者所为之行为。 ||

第40条（事实调查等）

第1款：发现以下违规行为或 | 施行令中无单独具体化条款。 | 无直接对应的指南。

收到举报/投诉时，可下达 | 调查程序遵循《行政调查基本法》。 |

资料提交命令或进行现场调查。 ||

针对条款：第31条第2款/第3款、 ||

第32条第1款/第2款、 ||

第34条第1款。 ||

第2款：进入营业场所进行调查。 ||

第3款：认定违规时，可下达 ||

中止命令或纠正命令。 ||

第42条（罚则）

违反第7条第9款 | 施行令中无单独具体化条款。 | 无直接对应的指南。

（泄露委员会职务上的秘密或 ||

用于目的之外）时， ||

处3年以下有期徒刑或 ||

3000万韩元以下罚金。 ||

第43条（行政罚款）

第1款：处3000万韩元以下行政罚款| 施行令第32条：行政罚款征收标准 | 无直接对应的指南。

第1项：违反第31条第1款事先告知 | （附表2） 。 |

规定 | 由科学技术信息通信部长官处以/征收。 |

第2项：未按第36条第1款指定 ||

韩国境内代理人 ||

第3项：未履行第40条第3款的 ||

纠正命令 ||

=====

=====

备注

1. 法案草案条款编号与现行法条款编号的差异

国会审查阶段的草案第15条（事先告知）、第16条（生成物标识）、第17条（深度伪造特例）对应现行法第31条（确保透明度义务）第1款、第2款、第3款。

草案第18条至第20条（高影响义务）对应现行法第33条（确认）与第34条（经营者责任）。

草案第21条至第22条（高性能义务）对应现行法第32条（确保安全性义务）。

草案第23条（影响评估）对应现行法第35条。

第2条、第3条、第4条、第36条、第40条至第43条在草案和现行法中的条款编号一致。

2. 指南的法律性质

五大指南是无法律约束力的行政指导文件。两项公告（《确保最前沿人工智能安全性公告》、《高影响人工智能经营者责任公告》）是依据法第32条第3款和第34条第2款制定的公告，具有等同于法规命令的约束力。

3. 过渡整改期

行政罚款的处罚自施行之日（2026年1月22日）起至少宽限一年以上。但发生人身伤亡事故、大规模侵犯人权等重大事项除外。

4. 视为履行其他法令（第34条第3款、施行令第27条第5款）

根据《个人信息保护法》、《医疗器械法》、《信用信息法》等现有法令履行同等措施的经营者，视为已履行第34条第1款的相应措施。具体对应关系已在施行令附表1中予以明确。

5. 来源

条款编号基于截至2026年7月21日国家法令信息中心（law.go.kr）的现行法令进行了核对。

《基本法》法令ID 014820（现行MST 282791），《施行令》法令ID 015032（2026年7月20日公布修订版，现行MST 288145）。

附录 C. 经营者类型分类清单

基于韩国《人工智能基本法》（以法律第20676号制定，法律第21311号修改为准）及《施行令》（以总统令第36053号制定，反映2026年7月21日施行的修改内容为准）。截至2026年7月21日。各项目均在括号内标注了依据条款。包含四项标准的韩国境内代理人指定要求规定在《施行令》第29条第1款中。五大指南是科学技术信息通信部于2026年1月22日发布的确保透明性、确保安全性、高影响人工智能判定、经营者职责及影响评估指南。

1. 开发经营者清单

☐ 确认本公司产品或服务是否符合人工智能系统的定义（《基本法》第2条第2项）

☐ 检查是否满足通过自主性、适应性及推理得出结果的要求（《基本法》第2条第2项）

☐ 确认基于规则的自动化（RPA、宏）不属于人工智能系统（《基本法》第2条第2项）

☐ 确认本公司是否属于开发经营者：指开发并提供人工智能的人（《基本法》第2条第7项a目）

☐ 检查在对外部模型进行微调并作为新服务提供时，是否有可能被归类为开发经营者（《基本法》第2条第7项a目）

☐ 确认是否属于生成式人工智能：是否为生成文本、声音、图像、视频等结果的系统（《基本法》第2条第5项）

☐ 确认是否属于高影响人工智能，比对11大领域（能源、饮用水、医疗卫生、医疗器械、核能、生物识别刑事侦查、招聘及贷款审查等权利义务判定、交通、公共服务、学生评估、总统令新增领域）（《基本法》第2条第4项a目至k目）

[] 实施高影响人工智能两阶段重大性判定：评估是否对生命、人身安全及基本权利产生重大影响（《施行令》第25条第2款、高影响人工智能判定指南）

[] 以书面形式记录高影响人工智能判定依据（《施行令》第25条）

[] 确认是否属于高性能人工智能：累计训练计算量是否达到或超过10的26次方FLOP（《基本法》第32条、《施行令》第24条）

[] 在提供高影响人工智能或生成式人工智能产品或服务时，建立向使用者事先告知其由人工智能运行这一事实的体系：告知时间、方式及内容（《基本法》第31条第1款、《施行令》第23条第1款）

[] 在生成式人工智能生成物上附带可见标识（《基本法》第31条第2款、《施行令》第23条第2款）

[] 在生成式人工智能生成物中插入机器可读标识（水印等）（《基本法》第31条第2款、《施行令》第23条第2款）

[] 对属于深度伪造的生成物另行附带可见标识，仅凭不可见水印是不够的（《基本法》第31条第3款、《施行令》第23条第3款）

[] 审查在明显体现其为人工智能或仅用于内部业务时，是否适用标识例外（《施行令》第23条第4款）

[] 属于高影响人工智能时，制定并运行风险管理方案：风险识别、分析、评估、处理及改进程序（《基本法》第34条、《高影响人工智能经营者职责公告》）

[] 属于高影响人工智能时，在办公室、营业场所或官方网站上公布主要标准、训练数据概要、使用者保护方案及管理负责人信息（《基本法》第34条、《施行令》第27条第1款、《高影响人工智能经营者职责公告》）

☐ 属于高影响人工智能时，建立应对受影响者说明要求的程序：说明用于得出最终结果的主要标准和原理（《基本法》第3条第2款、《基本法》第34条）

☐ 属于高影响人工智能时，建立人工干预体系：允许人工干预或推翻人工智能决定的机制（《基本法》第34条第1款第4项）

☐ 建立履行依据文件保存5年的体系（《基本法》第34条、《施行令》第27条第2款）

☐ 属于高性能人工智能时，实施风险识别、评估及缓解措施，并将履行结果提交至科学技术信息通信部（《基本法》第32条第2款、《施行令》第24条）

☐ 自主实施基本权利影响评估：事前准备、正式评估（编写风险场景及分析基本权利影响）、事后措施（《基本法》第35条、《施行令》第28条、影响评估指南）

☐ 检查训练用数据是否侵犯著作权并记录来源（《著作权法》）

☐ 确认在训练过程中收集的个人信息的处理合法性（《个人信息保护法》）

☐ 整理将向使用经营者提供的信息：人工智能系统的特征、局限性、使用条件及风险因素（《基本法》第34条、经营者职责指南）

☐ 确认根据现行法令（《医疗器械法》、《信用信息法》、《电气通信事业法》等）履行了同等措施时，是否属于被视为履行的对象（《基本法》第34条第3款、《施行令》第27条第5款、附表1）

☐ 确认五大指南并检查是否已将其反映在内部程序中（科学技术信息通信部指南）

2. 使用经营者清单

[] 确认本公司是否属于使用经营者：指使用开发经营者提供的人工智能来提供产品或服务者（《基本法》第2条第7项b目）

[] 检查与使用者（《基本法》第2条第8项）的区别：若仅获取人工智能并用于其本职工作，则属于使用者，不属于义务对象（《基本法》第2条第8项）

[] 确认是否已从开发经营者处书面获取关于人工智能系统的特征、局限性及使用条件的信息（《基本法》第34条、经营者职责指南）

[] 结合本公司服务背景审查获取的信息（经营者职责指南）

[] 若是使用高影响人工智能或生成式人工智能的服务，应向使用者事先告知其基于人工智能运行的事实（《基本法》第31条第1款、《施行令》第23条第1款）

[] 通过本公司服务提供生成式人工智能生成物时，履行标识义务（《基本法》第31条第2款、《施行令》第23条第2款）

[] 检查本公司服务是否属于高影响人工智能的11大领域（《基本法》第2条第4项a目至k目）

[] 属于高影响领域时，另行制定风险管理方案（《基本法》第34条）

[] 若开发经营者已履行风险管理、说明及使用者保护措施（《基本法》第34条第1款第1项至第3项），且本公司未进行重大功能变更，确认在该范围内是否满足被视为履行的要求，人工监督管理和文件化需另行履行（《基本法》第34条、《施行令》第27条第3款）

[] 认知若实质性变更了功能，则还需承担开发经营者的义务（《基本法》第2条第7项、经营者职责指南）

[] 建立应对受影响者说明要求的程序（《基本法》第3条第2款）

[] 使用高影响人工智能时，建立人工干预体系（《基本法》第34条第1款第4项）

[] 检查个人信息处理委托关系：区分与开发经营者的关系中属于委托人、受托人还是共同处理人（《个人信息保护法》第26条）

[] 建立使用者投诉受理及损害救济程序（《基本法》第3条）

[] 在服务条款中反映人工智能使用事实及其局限性（确保透明性指南）

[] 建立履行依据文件保存5年的体系（《基本法》第34条、《施行令》第27条第2款）

[] 确认根据现行法令履行了同等措施时，是否属于被视为履行的对象（《基本法》第34条第3款、《施行令》第27条第5款、附表1）

[] 确认五大指南并检查是否已将其反映在内部程序中（科学技术信息通信部指南）

3. 境外经营者清单

[] 确认是否适用域外管辖：本公司的人工智能服务是否对韩国市场或韩国使用者产生影响（《基本法》第4条第1款）

[] 检查提供韩语服务、韩元结算、面向韩国消费者进行营销等适用域外管辖的关联要素（《基本法》第4条第1款）

[] 确认在韩国境内是否有住所或营业场所：若有，则免除指定韩国境内代理人的义务（《基本法》第36条第1款）

[] 设有韩国子公司时，检查人工智能服务合同当事方是韩国法人还是境外总部（《基本法》第36条）

[] 韩国境内代理人指定标准1：上年度销售额达到或超过1万亿韩元（《施行令》第29条第1款第1项）

[] 韩国境内代理人指定标准2：人工智能服务部门上年度销售额达到或超过100亿韩元（《施行令》第29条第1款第2项）

[] 韩国境内代理人指定标准3：截至上年度末，前3个月日均韩国国内使用者人数达到或超过100万人（《施行令》第29条第1款第3项）

[] 韩国境内代理人指定标准4：曾有根据《基本法》第43条第1款第3项被处以行政罚款的记录时，无论规模大小，均产生指定义务（《施行令》第29条第1款第4项）

[] 符合上述四项标准中任何一项，即产生指定韩国境内代理人的义务（《施行令》第29条第1款）

[] 是否已书面指定韩国境内代理人并向科学技术信息通信部长官申报（《基本法》第36条第1款）

[] 确认韩国境内代理人是否为在韩国境内设有住所或营业场所的人（《基本法》第36条第2款）

[] 确认是否委托了韩国境内代理人的三项法定职责：提交高性能人工智能履行结果（《基本法》第32条第2款）、应对高影响人工智能确认要求（《基本法》第33条第1款）、支持履行高影响人工智能确保安全性及信赖性措施（《基本法》第34条第1款）（《基本法》第36条第1款）

[] 知悉韩国境内代理人的违法行为将归属于境外总部（《基本法》第36条第3款）

[] 用韩语向韩国使用者提供事先告知：由人工智能运行的事实、功能及局限性（《基本法》第31条第1款、《施行令》第23条第1款）

[] 将生成物标识义务应用于面向韩国使用者的服务（《基本法》第31条第2款、《施行令》第23条第2款）

[] 建立应对高影响人工智能确认要求的体系（《基本法》第33条）

[] 属于高性能人工智能时，建立提交确保安全性履行结果的体系（《基本法》第32条第2款）

[] 确认是否适用韩国《个人信息保护法》并建立合规体系（《个人信息保护法》）

[] 在面向韩国使用者的服务条款中反映人工智能使用事实、局限性及损害救济程序（《基本法》第31条、确保透明性指南）

[] 制定被处以行政罚款时的应对体系：未指定韩国境内代理人时，处以3000万韩元以下行政罚款（《基本法》第43条）

[] 确认是否属于仅用于国防或国家安全的例外适用对象：非专有（exclusively）的混合用途不适用例外（《基本法》第4条第2款、《施行令》第2条）

附录 D. 施行日程时间线

按时间顺序整理《人工智能基本法》及相关法律、规定的施行日程。

2022.3.1 中国《互联网信息服务算法推荐管理规定》施行

在中国境内运营算法推荐服务的经营者，须向网信部门办理算法类型和应用领域的备案手续。经营者由此承担了赋予使用者关闭算法推荐选择权的义务。

2023.8.15 中国《生成式人工智能服务管理暂行办法》施行

在中国境内提供生成式人工智能服务的经营者，须履行算法备案与内容安全保障义务，这也包括通过API提供服务的情形。

2024.12.26 韩国《人工智能基本法》在国会全体会议上通过（第20676号法律）

在场264人中，260人赞成。这是继欧盟之后，全球制定的第二部全面的人工智能一般法。虽然经营者目前尚未产生即时性的法律义务，但在公布后1年的宽限期内，已是需要着手准备的节点。

2025.1.21 《人工智能基本法》公布

正式名称为《关于促进人工智能发展及构建信任基础等的基本法》。确定自公布之日起1年后的2026.1.22起施行。经营者应开始排查自身服务是否属于该法的适用对象。

2025.2.2 欧盟《人工智能法案》禁止性实践条款开始适用

社会评分、实时远程生物识别等禁止性实践条款（第5条）以及人工智能素养义务自当日起开始适用。向欧盟 market/市场提供服务的韩国经营者也属于适用对象。

2025.8.2 欧盟《人工智能法案》通用人工智能（GPAI）义务开始适用

通用人工智能模型提供方编制技术文件、制定版权政策、公开训练数据摘要等义务自当日起开始适用。

2025.11.12 《施行令》（第36053号总统令）立法预告

进行了为期40天（截至12.22）的立法预告。期间面向产业界、学术界、市民团体等开展了70余次意见征集。高影响人工智能判断标准、透明度义务履行方法、韩国境内代理人指定标准等实务细节开始确定。

2026.1.20 《人工智能基本法》修订法公布（第21311号法律）

将国家人工智能委员会改组为国家人工智能战略委员会，并增加了促进公共部门引入人工智能、设立人工智能研究所的依据以及保障弱势群体可访问性的条款。这为拟进军B2G（面向政府供应）市场的经营者带来了全新的机遇。

2026.1.21 《施行令》公布（第36053号总统令）

正文共由33条组成。确定了11个高影响人工智能领域、高性能人工智能算力标准（10的26次方FLOPs以上）、韩国境内代理人指定要件（全球销售额达1万亿韩元以上、韩国境内人工智能销售额达100亿韩元以上、日均韩国境内使用者达100万人以上，满足其一即可）以及透明度履行方法。

2026.1.22 《人工智能基本法》首次施行

事先告知义务、生成物标识义务、高影响人工智能风险管理、高性能人工智能确保安全性、境外经营者指定韩国境内代理人义务正式生效。同时发布了五大指南（透明度、安全性、高影响判断、经营者职责、影响评估）。《人工智能基本法》支持服务台正式运行。对于行政处罚款的征收，给予了至少1年的引导期。

2026.5.14 美国科罗拉多州通过替代人工智能法案（SB 26-189）

2024年制定的科罗拉多州《人工智能消费者保护法》（SB 24-205）原定于2026.2.1起施行，但通过2025年85月州议会特别会议立法，施行日期被推迟至2026.6.30。而在施行前的今日，该法被以公开自动化决策技术（ADMT）和保护

消费者权利为核心的新法案所替代。新法案将于2027.1.1起施行。向科罗拉多州提供服务的韩国企业应以新法案为准进行准备。

2026.7.20 修订后的《施行令》公布（于次日施行）

其包含了将经科学技术信息通信部长官确认的人工智能产品和服务列为政府采购优先考量对象的确认制度（《施行令》第15条第4款、第5款），并在指定韩国境内代理人的标准中增加了行政罚款处罚记录（《施行令》第29条第1款第4项），使指定标准变为了四项。

2026.7.21 《人工智能基本法》修订法施行

国家机关等优先考量经确认的人工智能产品和服务的义务（第16条第3款）及采购经办公务员的免责条款（第16条第4款）正式生效。人工智能研究所的设立程序和向弱势群体提供费用资助的依据得以具体化。弱势群体的范围扩大至职业中断女性、求职者、非首都圈中小企业在职人员以及农渔业人员。

2026.8.2 欧盟《人工智能法案》一般性条款开始适用：公开聊天机器人身份、深度伪造标注、通用人工智能执行权限

透明度义务（第50条）开始适用，人工智能聊天机器人须向使用者公开其正与人工智能进行对话的事实，深度伪造内容也被赋予了标注义务。针对通用人工智能的执行权限和制裁措施自当日起开始启动。向欧盟市场提供服务的韩国经营者也属于适用对象。

2027.12.2 欧盟《人工智能法案》独立型高风险人工智能（附件三）义务生效（因《数字综合法案》修订推迟）

原定于2026.8.2起施行，但因2026年5月理事会与欧洲议会达成政治共识的《数字综合法案》修订，被推迟至2027.12.2。生物识别、招聘、贷款审批、教育评估、执法等领域的独立型高风险人工智能系统将被施加符合性评估、风险管理和文档化义务。违规时，将面临最高3500万欧元或全球销售额7%的罚款。

2028.8.2 欧盟《人工智能法案》产品嵌入型高风险人工智能（附件一）义务生效（因《数字综合法案》修订推迟）

对于医疗器械、机械、玩具、航空设备等适用欧盟产品安全法规的产品中嵌入人工智能的情形，将被施加第三方符合性评估义务。这原定于2027.8.2起施行，但因同样的《数字综合法案》修订而被推迟。向欧盟出口此类产品的韩国制造企业须在此时间节点前建立符合性评估体系。

附录 E. 风险管理方案及影响评估文件格式样本

本附录是高影响人工智能经营者必须制定的三项核心文件的格式样本。以虚拟的AI信用评估系统（以下简称“信用评分AI”）为例，添加了填写示例。在实务中，可根据本公司系统进行修改后使用。

格式 1. 风险管理方案（《人工智能基本法》第34条第1款第1项）

A. 基本信息

经营者名称：(株)金融科技解决方案

AI系统名称：信用评分AI v2.3

分类：高影响人工智能（第2条第4项G目：对招聘、贷款审查等个人权利及义务关系产生重大影响的判断或评估）

制定日期：2026年8月1日

制定者：AI伦理组 金OO组长

最终批准者：代表理事 朴OO

文件版本：1.0

下次预定更新日期：2027年2月1日（以6个月为周期）

B. 系统概述

（1）功能说明

信用评分AI是接收个人信用相关数据后，推荐是否批准贷款及利率水平的系统。供银行贷款审查负责人员作为参考资料使用。

（2）使用目的

在个人信用贷款审查过程中，计算信用等级、还款能力及违约概率，以建议是否批准贷款及适当的利率区间。

(3) 目标使用者

12家合作金融机构的贷款审查负责职员（约340名）

(4) 预计使用规模

月均贷款审查件数约为15,000件。每年在约180,000件的决策中被用作参考资料。

(5) 训练数据概述

自2018年至2025年积累的约320万件已完成非识别化处理的贷款履历数据。并入韩国信用信息院提供的统计数据。个人信息处理依据《信用信息法》第32条及《个人信息保护法》第28条之2（假名信息特例）进行。

C. 风险识别

(1) 歧视风险

已识别风险：可能发生因性别、年龄、居住地区导致的贷款批准率偏差

严重度：高

发生可能性：中

依据：训练数据中可能反映了过往贷款惯例的偏见。在2025年内部审计中曾发现特定地区居住者的贷款拒绝率具有统计学上显著偏高的模式。

(2) 隐私风险

已识别风险：在计算信用等级过程中，敏感信息（医疗费支出模式、向宗教团体捐赠等）被间接反映的可能性

严重度：中

发生可能性：低

依据：虽未在输入变量中直接使用敏感信息，但消费模式数据可能会作为替代变量发挥作用。

（3）安全风险

已识别风险：系统发生错误时，因批准不合格贷款而导致金融损失

严重度：中

发生可能性：低

依据：系统输出为参考资料，最终决定由负责人员执行。然而，由于负责人员完全听从系统推荐的比例高达87%，实际上接近于自动化决策。

（4）可解释性风险

已识别风险：拒绝贷款时，使用者无法理解拒绝事由的情况

严重度：高

发生可能性：高

依据：现行模型基于XGBoost，使用了数百个变量的非线性组合，难以向使用者解释个别变量的贡献度。

D. 风险缓解措施

（1）技术性措施

公平性测试：每季度一次，测量按性别/年龄/地区划分的批准率偏差。若偏差超过5个百分点，则重新训练模型或应用校正权重。

可解释性工具：计算SHAP (SHapley Additive exPlanations) 值，针对拒绝件提供排名前3位的影响因素。

异常探测：按日为单位监控批准率、平均利率及拒绝事由分布。若与上月相比偏离10%以上，则自动触发警报。

(2) 组织性措施

AI伦理组（3人）编制季度报告并向合规监察员报告。

外部审计：每年1次，委托独立的第三方机构验证公平性及准确度。

内部教育：以贷款审查负责人员为对象，每年实施2次AI系统局限性教育。

(3) 监控计划

实时仪表板：实时显示批准率、拒绝率、平均处理时间及提出异议件数。

事故报告体系：在使用者投诉中，将与AI相关的案件单独分类以分析月度趋势。

模型性能监视：每季度测量AUC (Area Under Curve) 。若降至0.75以下，则触发重新训练。

E. 人工干预体系

(1) 需要人工干预的决策类型

全数贷款拒绝件，适用最高利率区间（年利率15%以上）件，受理提出异议件

(2) 干预方法

由贷款审查负责人员确认AI推荐结果及SHAP解释后，做出最终批准或拒绝的决定。若推翻AI推荐，则将推翻事由记录于系统中。

(3) 负责人员

各合作金融机构贷款审查组长及以上职级。推翻权限仅限于组长级。

(4) 提出异议程序

收到拒绝贷款通知的使用者，可在14日内通过相应金融机构客户中心提出异议。提出异议件不经过AI系统，由审查员直接重新审查。

F. 解释方案

(1) 针对使用者的解释范围

拒绝贷款时：用自然语言告知对拒绝产生影响的排名前3位的因素。例：“最近6个月内新增贷款件数达到3件及以上”，“现有贷款还款额占年收入比例超过40%”

批准贷款时：告知对计算适用利率产生影响的主要因素2项。

(2) 解释方法

书面通知（邮件或电子邮件）+ 移动应用程序内的详细查询界面。

(3) 解释局限性告知

在解释文下方记载基于SHAP值的解释属于近似值，无法完全反映个别变量之间的相互作用。

G. 文件管理

保管期限：5年（施行令第27条第2款的法定保管期限，包括电子化保管）

更新周期：6个月（在模型更新、法令变更、发生事故时随时更新）

负责部门：AI伦理组

保管方法：包含版本管理在内保存于公司内部文件管理系统（DMS）。外部审计时可供阅览。

格式 2. 影响评估报告书（《人工智能基本法》第35条）

A. 评估概述

评估对象：信用评分AI v2.3

评估期间：2026年6月1日 ~ 2026年7月15日（45天）

评估执行者：AI伦理组 + 外部咨询委员2名（法学教授1名，数据科学教授1名）

评估目的：依据《人工智能基本法》第35条实施自主性影响评估

B. 社会影响分析

（1）积极影响

将贷款审查所需时间从平均3个工作日缩短至4小时。因审查标准一致性提升，减少了负责人员之间的偏差。扩大了小额贷款（500万韩元及以下）的可及性：引入AI后，小额贷款批准件数增加了22%。

（2）消极影响

自雇人士、自由职业者等非正规收入者的信用评估准确度低于正式员工。因高龄使用者（65岁及以上）难以访问数字渠道，可能被排除在基于AI的审查获益之外。随着系统依赖度加深，存在对负责人员独立审查能力弱化的担忧。

（3）受影响群体

直接影响：每年约18万名贷款申请人

间接影响：贷款申请人的家庭成员、保证人

C. 基本权利影响分析

(1) 隐私（个人信息自我决定权）

虽通过假名信息处理无法直接识别，但消费模式分析可能属于间接的用户画像。已单独实施个人信息影响评估（《个人信息保护法》第33条）并获得合格判定（2026年5月）。

(2) 禁止歧视（平等权）

公平性测试结果显示，性别间批准率差异为2.1个百分点，各年龄段间差异为3.8个百分点，均在标准值（5个百分点）以内。然而，特定地区（农渔村）居住者的批准率比城市居住者低7.2个百分点，正在施行校正措施。

(3) 听取解释的权利

随着基于SHAP的解释系统的引入，提供排名前3位的拒绝事由因素。但是，在使用者理解度调查中，回答“已充分理解解释”的比例仅停留在43%。正在进行解释文改进工作。

(4) 公平程序（正当程序）

已完成制定提出异议程序。2026年上半年受理提出异议件数47件中，有12件（25.5%）推翻原审定。

D. 环境影响分析

每次训练电力消耗量：约320 kWh（8台GPU服务器，72小时）

年度推理电力消耗量：约1,200 kWh

碳排放折算：每年约0.54吨 CO2（以韩国电力碳排放系数为准）

比较：与执行相同业务的人力相比，能源消耗显著较低，但在模型重新训练频率增加时，需要考虑累积环境负荷。

E. 综合评估及改进计划

综合等级：有条件合格

3项主要改进课题：

课题1. 农渔村居住者批准率偏差校正（目标：5个百分点以内，期限：2026年12月）

课题2. 改进使用者解释理解度（目标：“充分理解”达到60%及以上，期限：2027年3月）

课题3. 非正规收入者评估模型高度化（目标：自雇人士准确度AUC达到0.80及以上，期限：2027年6月）

F. 利益相关者征求意见结果

（1）内部利益相关者

对贷款审查负责人员进行问卷调查（127名回应者）：“对AI推荐的准确度感到满意”达78%，“曾推翻过AI推荐”达34%，“在推翻时感到有负担”达21%

（2）外部利益相关者

消费者团体座谈会（2026年6月20日，3个团体出席）：指出解释缺乏具体性，要求加强宣传提出异议程序

金融监督院事前咨询（2026年7月3日）：建议将公平性测试周期从季度缩短为月度

(3) 反映结果

消费者团体意见：反映在解释文改进工作中（改进课题2）

金融监督院建议：预计自2026年第4季度起转换为月度公平性测试

格式 3. 安全性检查日志（《人工智能基本法》第34条第1款第1项、第5项）

A. 日志格式

检查时间 检查项目 检查结果 采取措施 负责人员

- 2026.08.01 09:00 模型推理准确度（AUC） 0.82 正常范围（标准：0.75及以上） 金OO
- 2026.08.01 09:00 单日批准率 67.3% 较上月（65.8%）在正常范围内 金OO
- 2026.08.01 09:00 平均响应时间 1.2秒 在标准（3秒）以内 金OO
- 2026.08.01 09:00 异常交易探测 0件 不适用 金OO
- 2026.08.05 14:30 性别批准率偏差 2.3%p 在标准（5%p）以内 朴OO
- 2026.08.05 14:30 按年龄段批准率偏差 4.1%p 在标准（5%p）以内，进入注意区间 朴OO
- 2026.08.05 14:30 按地区批准率偏差 6.8%p 超过标准，采取应用校正权重措施 朴OO
- 2026.08.12 10:00 系统可用性 99.7% 满足标准（99.5%） 李OO
- 2026.08.12 10:00 受理提出异议件数 8件 较上月（6件）小幅增加，分析原因 李OO
- 2026.08.15 16:00 紧急检查：发生推理延迟 响应8.7秒 GPU服务器2号节点超载，分散流量 金OO
- 2026.08.15 17:30 紧急检查：采取措施后确认 响应1.4秒 确认恢复正常 金OO

2026.08.19 09:00 模型漂移检查 KL 0.03 在标准（0.1）以内，正常 朴OO

2026.08.26 09:00 月度综合检查 合格 已完成制定8月份综合报告书 金OO

B. 检查周期标准

每日检查：推理准确度、批准率、响应时间、异常交易

每周检查：公平性指标（性别、年龄、地区）、提出异议现状

每月检查：模型漂移、系统可用性、综合报告书

随时检查：系统出现故障、使用者投诉激增、外部环境变化（法令修订、经济指标骤变）时

C. 保管及管理

保管期限：5年（施行令第27条第2款的法定保管期限，包括电子化保管）

保管格式：电子日志文件（CSV或JSON）+ 月度综合报告书（PDF）

访问权限：AI伦理组、合规监察员、外部审计人员

备注：紧急检查件在发生后24小时内向AI伦理组长口头报告，并在48小时内制定书面报告书。

D. 事故报告格式（发生重大事故时）

事故时间：

发现经过：

影响范围（推定受害件数）：

紧急措施内容：

原因分析（第一轮）：

防止再次发生对策：

报告人：

报告时间：

填写示例：

事故时间：2026年9月3日 11:20

发现经过：在监控仪表板上探测到批准率较前一天暴跌30个百分点的异常模式

影响范围：对该时段处理的约120件造成影响

紧急措施内容：11:35 暂停AI系统推荐功能，转换为人工审查。12:00 着手掌握原因。

原因分析（第一轮）：前一日夜间批处理更新时发生变量映射错误。收入变量与负债变量的列顺序发生颠倒。

防止再次发生对策：批处理更新后引入自动回归测试。增加变量映射校验和验证程序。

报告人：AI伦理组 金OO

报告时间：2026年9月3日 15:00

Support This AI Library Book

If this book was useful, you can support future translations, public editions, and open AI knowledge projects through PayPal.



PayPal

[**https://paypal.me/kimkjj**](https://paypal.me/kimkjj)

Scan the QR code or open the link above.