人工智能系统的规制

[德]托马斯·维施迈尔 著　马　可 译　赵精武 校*

内容提要：人工智能系统的大规模应用向法律提出了巨大挑战。由于目前智能系统的发展尚未达到人类智能的水平，人工智能系统的不可预测性和不可控性将会引发一系列问题，因此有必要对人工智能系统应用模式进行监管。目前人工智能系统有三个核心技术要素，即数据、算法和人工干预。这三个要素对于监管原则的确立有着锚定作用，同时也是监管制度实施的着力点和立法调整的规制对象。监管人工智能系统的最大障碍是它的决策过程难以被相对人理解，这意味着人工智能系统需要在决策过程中保持高透明度，以此确保可理解性。但高透明度是难以通过事后的决策解释来实现的，而应当是通过决策时系统提供决定理由来实现的。因此有必要建立一个基于智能系统决策的理由和监管架构，在架构中将机器决策纳入法律程序中。这样的架构可以使得决定方和相对方充分交换意见，使得个人真正成为数字化秩序的主体，而不是客体。

关键词：人工智能　智能系统　透明度　法律监管

DOI：10.19350/j.cnki.fzsh.2021.05.010

一、走向开放：人工智能对法律的挑战

在信息时代，在人们决定、解释某事，或是实施某个行为的过程中，机器的身影频频现身。具备学习能力的信息系统早已应用于医疗、金融等领域。网络空间中的智能算法也在根据用户搜索时所透露的信息，为用户量身定制产品推荐和相关信息。随着世界从互联网时代走向物联网时代，人工智能技术也愈加深刻地融入人们日常生活中：各种传感器在住所、交通工具以及身体佩戴的机器设备上随处可见，实时收集数据并将数据传输至云端。云端的智能系统则根据这些数据完成人类指定的任务。这些智能系统能够在现实世界中操纵机器、移动物体。它们正在迅速地超越智能机器人的阶段，向着更高阶的智能发展。因为它们具备这些特定的能力，智能系统不单单应用于私人领域，同样也被应用于满足公共服务需求的公共领域。如警方应用预测警务系统来预测犯罪行为。而在具体的国家行政监管中，特别是在金融市场监管和税务征收管理等领域，智能技术都有很大的用途。

目前智能系统还远远不能与人类智慧相提并论。故而关于智能系统的争论焦点并不是未来才可期的“超级智能”，而是如何进一步发展和应用具有自主学习能力的智能系统，以及这种系统对于大数

* 托马斯·维施迈尔（Thomas Wischmeyer），德国比勒费尔德大学（Universität Bielefeld）助理教授；马可，北京大学法学院博士研究生；赵精武，北京航空航天大学法学院助理教授。本文原文发表于德国《公法档案》杂志第 143 期（2018），第 1–66 页，本译文系中文简易版，不带注释并对原文进行了删节。本文系中国科协 2020 年“高端科技创新智库项目——政府数据安全与共享的现代治理研究”（项目编号：20200608CG080713）阶段性成果。

据进行开发和利用的能力。将“一般的人工智能”从“应用型人工智能（狭义上的人工智能或弱人工智能）”中分离出来的意义也是如此。但即便是目前的“弱人工智能”或“应用型人工智能”，也会对政治、社会和法律提出新挑战，特别是当智能系统被整合进入具有巨大社会影响、有造成损害可能性的决策过程时，其产生的问题尤为突出。这种情况愈演愈烈，尤其是在金融市场或安全管理等重要领域中，智能系统的不可预测性、不可控性很可能引发严重后果。人与机器之间紧密的相互作用性、机器学习技术的广泛适用性以及人与机器的共生性更加剧了这一现象。因此，有必要明确，系统越独立，其行为就越有可能摆脱人类的控制；换言之，技术已经从人类的工具变成了规范规制的直接对象。

基于前述社会问题的观察，本文意欲解决两个核心问题：首先是澄清目前智能系统规制问题中的核心议题，即是否有必要通过立法途径实施对智能系统进行有序且系统的监管活动。基于现行的宪法性文本和基本法律，可以归纳出智能系统设计和应用应当遵循的六项监管原则。在此基础上，本文进一步分析目前智能系统规制的前沿问题，即如何规制智能系统的不透明性。智能系统的决策机制是一个“黑箱子”，即难以验证哪些因素影响并导致了特定的机器决策结果。虽然大众都对于智能系统有着一定的了解，但立法层面所追求的智能系统透明性标准已经超出技术安全问题的实际范畴。复杂的智能系统做出决策作为一种“黑箱子”现象并不构成实质的法律问题，因为人类做出决定的过程相对于他人而言其实也是“黑箱子”，从规范的角度看，机器决策与人类决策并无实质差异。因此，相比于机器决策的透明性和可追溯性实现问题，更重要的是，需要通过适当而完备的法律程序，如下文所言的推理程序，来规制机器决策的过程。虽然人们对于人工智能的发展有着不切实际的幻想，但是技术工作者都能认清目前人工智能技术的发展水平和局限。因此，在进入法律上的讨论前，必须对智能系统的一些前提条件进行澄清：如今的智能系统能力到底如何？它与传统的自动化决策系统有何不同？它的合法性边界在哪里？

二、智能系统的监管原则

技术并不能告诉人们该做什么，但它能够提高人们对于规范性问题的认识，并推动关于如何制定、实施法律规范和标准的讨论。从监管的角度来看，在规制智能系统时应当考虑下列问题：哪些领域的法律规范已经存在？立法者有无必要对这些法律规范进行修订或调整？或者新技术所带来的新问题是否已经无法通过现有监管方式予以预防和规制？

监管的基本原则在被实施并被具体化到宪法基本价值的过程中，立法者应当要留有一定的余地。在个案中，监管的基本原则之间会存在紧张关系，立法者必须在基本原则之间寻求动态平衡。在监管规则的配套程序机制下，根据主体公私属性的差异应当配置不同的监管制度。此外，配套程序机制也应当根据智能系统的应用领域进行不同的制度安排。最后，开发和使用智能系统的人也可以主张宪法层面的保护，进而不受特定监管规则的制约，这也会进一步推动前述法律程序差异化的处理，下文不对此作出论述。因此，对于警察法等特定领域的智能系统监管规则，重要的是要符合公私法领域的一般性规范、标准。

（一）智能系统的决策过程应当做到“可视化”

每种技术中蕴含的价值判断会在应用环节中持续体现自己的影响力，并限制技术在特定领域的应用路径，最终在该项技术的逐步发展、完善中成为应用纵深化的实践基础。例如在网络系统的搭设和形成中，软硬件与系统结构的具体标准在一开始就包含在网络行为管理设备的条形码中。算法和智能系统也并不是如一般社会认知般是纯粹技术性的、人造的和中立性的，而是充满耦合性，并与其提出时的语境紧密交织在一起。在早期对核心算法的研究中，这个观点被反复强调，特别是在

系统设计阶段，系统的决策早已被界定在一定的范围内。如果我们没有准确地认知技术的力量以及技术产生的背景和后果，将会难以准确地批判技术的现实问题，进一步阻碍公众的自主行为并限制他们对于民主自治空间的感知。因此，监管或技术的可视化，通常被理解为社会技术与监管技术设计的重要目标。在智能系统中，这一目标的重要性主要体现为：首先，智能系统可以动态响应用户的行为，相较于其他的技术方案而言更大程度上参与了社会决策。其次，由于其广泛的适用性，智能系统的决策很少受到特定需求的影响。

应当强调的是，我们能够处理的仅限于可视化、数字化的数据流，然而，当我们忽视（智能系统存在促使社会公众进入“数字无意识”的）特性时，即忽略了它可以在不知不觉之中影响、确定人的行为时，就会诱发一系列社会问题。这种现象也可称为互联网的副作用——即Eli Pariser所提出的“过滤泡沫”[①]现象——这种现象已经在政治领域和社会中存在了相当长的一段时间。但即使智能系统从事其他事务——而不是保留和输出用户想要的信息，人们仍然会在“数字无意识”的影响下慢慢作出他们潜意识状态下的思考结果。但是，由于通常情况下用户们缺乏相关领域的知识和经验，他们并不了解具体的决策过程，他们无法将这些行为与监管要求或者法律责任建立起联系。如果无法让主体识别出这些现象和行为，就会导致用户权利在事实层面受到限制，索赔也会遇到障碍，且举证责任的实际难度远超以往。这种极不合理的社会现象有必要通过某种方式将智能系统的决策过程公开化和透明化。此时，从公法的权利义务关系的角度观察，法律所要求的可识别性标准顺势成为法治国原则的基本要求。

（二）智能系统应符合一定的质量标准

智能系统的发展和进步基于复杂的数学原理和技术方案，而这些恰恰是立法难以直接触及的领域。因此，除了自我学习系统设计所取得的巨大成功之外，正如Tay的例子，也会存在一些“挫折”。

如今，许多标榜自己为人工智能的应用程序事实上并没有达到智能系统的最低技术要求，同时还缺乏高质量的数据实现内部的迭代优化。在“向系统输入垃圾，输出的也是垃圾”的机理下，这些应用程序根本无法做到降低人工成本，更遑论潜在的安全漏洞威胁。此外，有的所谓智能系统数据库中的算法和得到的相关性是否实际具有预测能力，是否只能识别非生产性数据（也就是仅仅是人类的辅助工具），这些潜在的问题通常也没有得到足够的重视。如果系统或开发者没有关于特定领域因果关系的知识，或是忽略了潜在的重要因素，上述问题会体现得愈加明显。人们总是忽略了以下事实：在自然或社会里一些没有规律的场域中，人工智能很难找到可靠的规律和因果关系。

这样的错误显然会影响智能系统的质量和可靠性，但即便是能够避免这样的错误，智能系统也不当然是完美的。智能系统预测准确性的高低，取决于所使用的预测模型和系统设计者开发时花费的精力。此外，由于智能系统的决策模型采用了很多统计学方法，因此会始终存在一定的“假阳性”和“假阴性”决策。如果将智能系统应用于法律领域（比如刑事指控），由于系统基本上只能进行相关性的论证，统计学层面的报告难以作为法律判断的根据。

立法者或可考虑在法律层面设置充分的技术标准，特别是涉及基本权利的敏感领域。如此一来，有助于辅助并以国家保证责任（Gewährleistungsverantwortung）的实现，成为民主、法治国和其他宪法价值的坚实基础，同时也给予人们合适的途径去救济自己的权利。在监管层面，应当根据具体情况精确定义具体的质量标准，或是确立能同步同期最高科技水平的技术开发设计流程与程序。

① 所谓的“过滤泡沫”，是指越来越智能化、定制化和关联化的网络会限制用户的世界观，用户仅能在智能系统中通过过滤和筛选的信息“狭隘”地观察现实生活，被困在“网络泡沫”中。

现行的部分法律已经初步地确立了最低的技术要求，如德国《税法通则》（Abgabenordnung）的第88条第5款第3项向税务机关使用的风险管理系统提出了最低的技术标准。但在大多数情况下，当机器决策存在缺陷或是机器作出超出程序限制的行为时，仍然缺乏足够的规范性判断标准，这比法律评价人类行为更为困难。

针对前述问题，可行的措施首先是在某些重要应用领域，面向关键的基础设施和系统模型确定法律层面的技术标准和质量标准。立法者也可以通过强化法定的注意义务、审慎标准，或重新分配责任、风险的方式来激励对高质量技术的投资。当然，立法者也应当避免过于苛刻的法律责任和标准妨碍创新技术的应用，最大程度保证技术可以满足市场的需要。此外，软件开发活动往往是有组织的、高度分工的，因此在开发过程中一定会对风险分配做出预先的安排，立法者应当在事前阶段作出合适的责任分配和风险分配。立法机构还可以通过立法活动加大对相关领域的投资来优化智能系统的决策质量。从这个角度讲，监管智能系统与监管其他具备复杂性和风险性的技术（如药品和纳米技术）具有同质性。

确定技术标准亦可帮助确立责任制度。判断某种智能系统行为或错误的决策行为应当适用何种具体制度关键在于该智能系统应用的特定领域和范围。如在Twitter机器人Tay案例中，当具体讨论人工智能操纵机器人作出不当输出时，即这些错误算法和深度学习造成了不良后果并可能产生法律责任，我们有必要审视和讨论几个非常重要的问题：谁来承担民事责任？如果涉及刑事领域，谁应当承担刑事责任？——这些问题显然有必要专门论证。

（三）智能系统与反歧视

智能系统的广泛应用也可能会导致歧视和社会结构性转变。诚然，研究人员和制造商正在使用智能系统来试图剔除人类决策中固有的任意性因素；当去除了主观偏好后，机器应仅根据客观标准做出决策。技术也可以屏蔽甚至是删除某些敏感的数据类型。但是，实证研究表明，实践中大部分智能系统的应用方式都有可能产生歧视效应。

因为智能系统常常应用于大规模用户的数据处理，所以智能系统自身的歧视性态度会显著而强烈地影响用户群体。在非个性化推荐模式下，用户主观偏好的分散完全利用统计学的“平均”方案予以“中和”，以此消除和避免针对特定群体的社会歧视。但在智能系统的影响下，人们原本多样化的决策行为将被单一的决策行为所取代。如果我们不假思考地接受“技术中立”立场，则会造成三个层面的社会风险：巩固对社会有危害的观念、歪曲客观观点和影响公众的自我认知。即便是智能系统只是应用于辅助人类决策的特定场景下，具有歧视倾向的智能系统仍会间接地影响人类的决策。但即便是预先过滤歧视性数据，也难以真正摆脱智能系统对最终决策结果的影响。比如，预测性警务系统在分析数据过程中发现特定宗教信仰和犯罪概率间存在统计学上的显著相关性，并依据这些数据提出建议，而这些建议一旦被人类决策者采纳，最终的结果只会是加剧人类决策者针对特定族群的“异常”管控。这种社会现象本质上构成了对立法、执法和司法活动的干预，换言之，相当于智能系统替代权力机关对特定族群实施特定行政行为。

智能系统或算法会造成决策偏差的原因是多方面的：歧视效应既可能是由于输入数据缺乏代表性和科学性，也有可能是在系统设计时就被植入人类偏见。当然，也有可能是输入端与输出端的数据集合之间确实存在客观的关联性。但当这种关联性与性别、种族等要素相关时，人们应当保持足够的警惕性。

在智能系统的设计过程中，很难避免潜在的歧视性因素。即使是先进的智能系统也依赖人的输入。智能系统设计过程中包括数据选择、算法设计和不同强度的检验训练。在这些环节中，人类的

主观价值会以潜移默化的方式植入系统中。不过，这种潜在的系统歧视也可以通过相应的缓和机制予以控制。例如，程序员可以禁止系统、传感器自动关联特定因素（如性别、种族），事前规避对某些特定类型数据（如肤色）的收集。但这种缓和机制的效果着实有限，在部分场合下难以控制系统通过技术间接歧视的情况。众所周知，一些明显的中性特征往往与一些敏感群体的特征相重叠。例如，预测警务系统一开始就忽略所涵盖特定的宗教相关数据，但是根据所给定的数据，仍然将犯罪的可能性与特定的地理区域相联系起来，而该区域居住的人群主要是某特定宗教的成员。这种潜在的歧视和偏见难以判断，也难以纠正。更重要的是，在法律视阈下，这种歧视和偏见的产生时间、刑事责任的认定期间等问题难以明确回答。

为了进一步减少这种情况，必须要在每个环节对系统结构进行技术层面和制度层面的细致安排。开发者和使用者应当谨慎选择机器学习的训练数据集合，也要谨慎审查对数据分类标准的设计和分类模型。不论系统自动决策的评价标准和结果如何，都应当注意维系个体权利和整体社会福利的平衡。

但如果上述的歧视影响很大程度上得到了解决或是补偿，是否意味着上述的类型化限制措施矫枉过正了？在人们完全依靠系统和技术来进行反歧视前，或者在智能系统由于内在的歧视倾向而受到怀疑之前，这些措施可以让我们意识到不同智能系统中的数据库之间的差别有多么大。在特定领域中，如医学诊断领域，预后系统是高度个性化的。从歧视法的角度来看，这几乎没有问题。不过，根据可用数据将个人视为特定群组组成部分的系统也存在问题。若政府希望使用人工智能系统的预测功能，实现在具体危险出现之前有效地分配风险和资源，那么，事前可以输入的数据质量和数量明显不能够满足预期目标。在没有输入数据对智能系统进行训练的情况下，危险事件本身就是数据的来源。此时，周期性的个性化数据收集要么是还未开始，要么就是还不合法。通常情况下，智能系统的数据库在初期仅有非常粗略的数据，如事件的类型、犯罪统计数据和一个地区的社会经济状况。

立法机构应对前述风险所能采取的有效措施极为有限，如果存在特定的充分合理原因，则应当尽早采取前述所提到的矫正措施。然而技术研究表明，由于系统结构原因，当缺少个性化信息进而导致系统可能会在可预见范围内被滥用时，应当放弃使用智能系统，或是为行政相对人提供额外的制度保障和救济渠道。这个议题也可以在智能系统的应用这个更大范围的议题内被讨论。但必须要说的是，通过收集更多的个人数据来扩大信息库，有助于决策的个性化，也就是说，上述对于智能系统应用的规范并不是在所有场景均可适用，其适用前提应当仅限于出于禁止歧视并保护数据之目的。

（四）智能系统使用过程中的数据保护和信息安全

智能系统的成功很大程度上取决于它们可以高效处理大量数据。不过，如果这些数据属于法律保护的个人数据，收集和处理行为显然应当遵守众所周知的数据保护和信息安全规则。大部分的法律规则和标准均重点关注此环节，且整个欧盟成员国国内法均在维系与《通用数据保护条例》（以下简称《条例》）标准的统一。在数据保护方面，对智能系统的监管已经相当完备了，智能系统在运作中出现的新型法律问题并不是数据保护法的问题，而是设计其他法律领域的治理问题。

下面将进行更深入的教义学分析，进而可以处理前述的数据保护和反歧视的法律适用问题。比如，从数据保护的积极面来看，有市场的智能系统对于受影响者的身份不感兴趣，因此对友好的隐私保护政策是持开放态度的。从反歧视的角度来看，放弃收集某些类型的数据往往是有好处的，但是，当系统缺乏指定数据却仍然做出决策的时候，就会产生前文提及的特殊问题，即“输入的数据存在问题使得结果失真，从而引发了本应当避免的损害”。显然，生搬硬套地限定收集数据范围无

助于问题的解决。相反，应当在权益平衡的大前提下兼顾相关主体的权利以及与决策结果关联的各种规范性文件。无论如何，在谈到有利于数据保护的措施时，立法者应该谨记保护所有弱势少数群体的隐私。欧盟法律中个人数据的广泛概念使得智能系统无法逃避法律的监管，其决策结果的实质影响也被限定在一定范围。数据保护法本身不能解决主体间的紧张关系，但是它可以通过强制性规范解决可能直接影响个人权益的部分问题。

大体上看，数据保护法限制权力机关和私人机构无限制地收集数据，限制了潜在的算法歧视风险，但客观上也限制了智能系统在其他场域充分发挥其能力和长处的机会——这可能会产生其他的法律后果。

（五）智能系统使用过程中存在的多种问题

应用智能系统的一大准则是扬长避短。智能系统的优势是可以应对大量数据和且可快速准确地决策、预测，发挥这一优势的前提则是要考虑特定应用领域本身的背景特点，如智能系统可以应用于金融市场的预测。但如果需要结合不同策略来解决问题，或是需要开辟新的信息来源、发展象征性理论，则须谨慎地选择智能系统应用方式，因为智能系统在解释和编写文本方面的能力是有限的，市场中常见的系统尚且无法就个案中的所有情况作出决定。所以，为了实现智能系统相对于人类决策者的时间成本优势，其应用场景通常限定于只处理部分标准化数据。

在部分应用场景中，法律会限制智能系统的应用，规范并不能完全涵摄系统处理数据的所有方法，即规范并不能与现实场景一一对应。这一社会问题的形成原因兼具事实层面和规范层面的因素。比如，尽管大数据技术已经相当成熟，但世界上大部分地区尚未被传感器覆盖；此外，以数据保护法为代表的新兴立法活动往往限制了出于私人或公共用途数据收集的可能性。这些情况都可能导致系统的功能不能符合预期。

在上述这种情况下，私主体如果仍然使用智能系统，就有可能遭到损害进而有损害赔偿请求权，或者说至少有得到赔偿的正当性基础。上述情况对于公权力机关还有更进一步的要求。在法治国的要求下，并不存在要求每项决定都完全的个性化考量。也就是说，行政活动就是公式化的，而很多时候“公式就是程序”。存在允许模式化的空间的前提下，智能系统就有发挥长处的空间。而在政府行为的应用场景之中，“粗数据”和粗略的描绘是不够的。特别是在安全法领域，教义学规范要求智能系统必须为行政相对人提供特定信息，只有提供信息才能为政府做出的特定干预措施补足正当性。因此，在进行制度设计的时候，必须始终注意要留有“诊断窗口（Diagnosefenster）”。现在，在设计智能系统的时候也要考虑这一点。

根据欧盟法律，智能系统的决策须考虑到个性化诉求是有充分根据的，特别是《条例》第 22 条第（3）项和《通用数据保护条例序言》（德文简写 ErwGr，英文为 Recitals of GDPR，以下简称《条例序言》）第 71 条 U 部分的第（2）项第 1 句提到了“‘人有权进行干预’作为可能的保护措施……并陈述自己的观点和对决定的争论”。对于政府行为的实施者，德国《联邦行政程序法》第 24 条款第 3 句与《社会法典》第 10 部分第 31a 条第 2 句对调查的原则进行了规定，主管部门在使用自动系统无法确定行政相对人的时候，有义务考虑对方的相关事实陈述和他们遭遇的实际情况。《税法通则》也在第 155 条第 4 款第 3 句规定了在自动化系统决定中遭受不利影响的人拥有主动选择手动处理他们事务的可能性手段。

这些例证说明了如何平衡制度运行中不同的利益和后果，也说明了如何应对个体个性化的诉求与规范稳定性、确定性之间的张力。对于未来，须考虑到智能系统不一定要超越人类的头脑作出决定。如果当前的系统仍然基于较粗糙的数据运行，那是因为系统设计者决定这样做，但是这种应用

现状并不能反映该技术的全部能力。数据保护法完全可以考虑针对基于精细数据训练的系统设置“数据具体程度”层面的限制。因此，干扰系统自我选择信息的措施实质上并不符合社会治理的真正需求，因为数据过少只会导致无法有效运作系统——但是从干预本身的理由来看，对数据收集的限制和禁止却在不断增加。

正如前述，智能系统的相关性和统计数据是不充分的，这并不足以证明因果关系。通常，这不会妨碍到智能系统的应用，即使没有令人满意的解释模型去揭示具体的因果关系，法律也允许它根据以往的经验数据来进行决策。但在特殊领域是不够的，如在刑事诉讼领域确定被告人是否有罪的时候。如果坚持认为决策不能基于统计的相关性，而必须基于对因果关系过程的细致重建，那智能系统的使用空间就很有限了。

（六）智能系统使用过程对于责任和职责的明确化

目前的规范有着相应的责任程序，其主要针对对象是立法者和行政人员。立法者和行政人员可以与智能系统的开发者、用户和其他利益相关者进行直接接触。根据前面提及的程序，立法者和行政人员可以要求开发者在设计系统时使其规范效果可见，并确保适当的技术质量水平，实现预防歧视、保护数据安全、注意行政相对人的个性化诉求等要求。该责任程序可以通过传统的法律话语体系，特别是法释义学的责任归属语境予以表达。但是，责任的准确分配终究是一项非常困难的工作，现今仍无法确定是否需要在新领域引入新的责任制度。

即使立法者的制度供给充足，智能系统依然可能出现错误并造成损害。损害的原因通常包括开发者设计不当、用户使用不当，以及数据统计中不可避免的“假阳性”和“假阴性”。此外，机器和人类都会面临同样的问题，即行为的负外部性，在任何情况下，行为都有可能产生责任问题。如今智能系统行为的负外部性正在挑战现有的规范，因为与传统的自动化系统不同，具备学习能力的智能系统的行为并不能完全由程序员决定。不具备学习能力的系统出现决策过程故障时，原则上可以归因于设计、使用环节或以其他方式与整个过程产生联系的法律实体行为，但（具有学习能力的）智能系统的归因路径由于原因、事件之间的关联仍难以明晰。

在这种背景下，新引入的责任归属制度是否需要进行根本性的改革仍是一个重要议题，同时也有其他关联问题，比如，是否可能有限度地承认智能系统的“数字法律人格”，使系统成为权利和义务的自主承担者；能否将其看作代理人或是履行辅助人。这些议题也可以与法哲学中的争论相结合，也就是职责（Verantwortung）和责任（Schuld）是否可以加于机器之上。这种讨论当然也是有意义的，因为智能系统是否具有意识是承认其权利能力或法人资格的基础。

从法律监管的角度来看，在此背景下须考虑三个要素：

第一，目前的系统设计工作仍然是由众多人类协同完成的。且到目前为止，这些系统不是自主运行的，而是由自然人或法人控制的。因此，每个智能信息系统的行为事实上都不过是人类行为的外壳，系统开发者、用户和利害关系人仍有可能继续被视为实际的参与者。智能系统责任的讨论绝不能忽视这些主体的行为。鉴于这一领域中的不同主体（公司）间的高度合作性（比如智能系统通常由无数可自由使用或交易的单独组件组成，而这些组件有各自的开发者），因此风险和责任的具体划分显然不是毫无意义的。

第二，必须考虑“人”这一法律概念的特殊性。诚然，《基本法》第1条第1款将权利能力给予了所有人，并将法律主体的基本概念与法律之外对于人格的确认联系起来。但从纯粹的法律角度来看，这只是一种技术上的操作，除了将授予《基本法》第1条第1款所设立的“自然”法律主体的权利资格之外，立法者还有其他立法技术来对智能系统进行界定。因此，立法者原则上可以自由

地将智能系统界定为新的法律主体类型，前提是这种界定方式有助于简化论证结构和归咎责任。当然，这样的处理方式也会带来许多后续问题，比如创新技术设计的责任归属、保险标的认定以及新型法律主体的责任财产问题。这种功能性考虑的目的而不是对智能系统性质进行形而上角度的探索，对于在法律责任体系中寻找智能系统的定位，并对它作出恰当的安排至关重要。

第三，法律责任的归属和分配方面仍存在很大的不确定性。在受害者层面，责任的准确分配是次要的，谁来承担损害赔偿责任才是核心。因此，即便事情很复杂，但只要责任主体并没有过大扩散和传递，那么，为了保证受害方的损害赔偿请求权，立法者一方面必须制定明确的责任规则，另一方面要确保可能索赔能够被强制执行。考虑到现有的责任法在处理传统软件和信息系统相关问题时已经趋于疲软，因而可能需要在立法层面作出大规模修订。此外，由于智能系统的复杂性和动态性，通常很难确定到底是哪种类型的系统行为导致了损害结果的发生。当然这些问题或可通过诸如严格责任等责任法制度来解决，但这些方案并没有从根本上解决关于智能系统法律规制的核心问题——系统行为过程的透明度。在分析这个问题之前，有必要明确的是，确保智能系统行为过程的透明度是主权者的职责。

（七）特别注意——主权者对智能系统的使用

从二十世纪四十年代起，人们就认为政府和国家对智能系统的应用不会落后于私营部门，政府会迅速应用智能系统来处理公共事务。从那时起，技术控制论作为新兴学科得到迅猛发展，这也激发了人们对于自动化政府和行政管理的新愿景。然而，行政自动化长期局限于用机械替代人类从事辅助类的管理活动，即便国家应用新技术的速度非常快，且现今没有国家可以离开计算机维持税收、社会保障及福利制度的运行，但这些自动化技术的应用形式仍停留于初级阶段，并不能引起行政管理活动形式和性质的根本变化，行政法学理论也就无须快速予以回应。

但是技术所代表的可能性改变了人们对于法律制度自动运行的态度。在自动化应用的早期阶段，为了将较为原始的技术引入行政管理，需要突破现行立法中实质性和程序性要求。如今这种现象已经不存在，人们已经拥有了功能更强大的计算机、更多数据和更优化的算法，“智能治理”这种全新的行政管理模式已经出现。前述的例子表明，这种可能已经在数据安全等敏感领域得到了广泛的应用。现在，智能系统支持下的司法活动和“机器人监管”正在一步步被实现。但目前技术的能力充其量仅能用于简单的涵摄。因此，智能技术的应用意义主要集中于事实判断层面，但在更复杂的合法性判断上意义仍然不大。

在目前关于法律政策的讨论中，人们正在努力为自动化行政奠定相应的法律基础。但与此同时，人们也应当重视法治国原则。立法者修改了《行政程序法》第 35 条 a 款的规定，尝试通过自动化的手段发布具有法律约束力的决策。在这种情况下，确定事实和涵摄外延均交由机器来完成。尽管《行政程序法》第 35 条 a 款的具体内容存在争议，但这一立法活动已经释放了一个信号，表明智能系统不再只扮演行政活动的辅助角色。即公民原则上能够独立发动一个行政程序，即使是在没有行政人员参与的情况下，最终也能得到一个具体行政行为式的回应。据此，可以预想的是，即将出现满足条件即能自行启动的智能系统。

但仅允许智能系统实施行政行为是不够的，相应的授权和权力边界也必须予以明确，这也是在整个行政智能系统决策过程中应当贯彻的指导原则。前述的监管原则只要被宪法辐射到，即对权利机关具有约束力。根据系统的特征和所涉及领域的特点，这些彼此存在张力的规则，可以限制、授权智能系统的应用。为维护宪法与法治国原则所确认的基本价值，也需要为智能系统的行政行为设置配套的程序性事项与实体制度，比如《税法通则》第 88 条第 5 款第 3 项——对财政主管部门应

用智能系统的条文。监管制度的理想效果是能够调和技术可行性和监管目标之间的紧张关系，并将系统的应用场景限制在特定的专业领域中。

在贯彻专业领域的智能系统监管原则的时候，人们应当意识到通过智能系统决策取代人类决策将会改变决策的制约因素，因此需要依据差异化要素区隔人类决策和机器决策，比如速度、一致性、精确性、灵活性等。在实务中，这种区别会产生外溢效应。因此，智能系统决策的低成本和决策的高一致性可能会使立法者面临下述情况：由于智能系统决策的固有特点，在智能系统行政尚未应用于日常生活领域时，人们不得不接受相对较高偏差率的“人力执法”情况。一旦当智能技术全面实施产生高昂的后续成本时（因为机器仍然会犯错，并且会导致大范围的错误），这会使得公民无法守法，导致法律规则受到侵蚀，此时立法者必须阻止自动执法、改变实体法或是为“有用的非法（brauchbare Illegalität）”创造空间。

三、透明化的智能系统：从解释权到分工建立、控制架构

前述原则可以指导国家预防、纠正应用智能系统的负面效果。立法者和行政部门必须采取制定智能系统的技术质量标准、采取预防措施、制定隐私和信息安全标准等措施，不过这些措施并不容易被实施。正如以往面对新技术的冲击一样，国家必须努力获得必要的监管技术和能力才能正确履行保障国家的责任，那么有效地监管智能系统开发并在适当情况下进行干预成为必然选择。

但这在智能系统的监管方式能够取得预期效果吗？从现有研究成果来看，人们普遍强调，研究清楚人工智能这个“黑箱子”的决策机制几乎是不可能的。事实上，系统功能越强大，系统的决策过程就越难被理解。但是，如果没有透明度，就很难正确认识和感知法律应当设置规范的实际范围。

关于透明度问题（Das Transparenzproblem）有着很多争论，但比起其他相关议题，这些争论鲜有涉及技术原理。智能系统使用的逻辑和技术是非常先进且复杂的，但这种复杂性仍不能与原子技术和纳米技术相提并论。但智能系统的具体决策过程对于平常人仍是不透明的，或者正如Jenna Burrell所言，“因为受决策影响的人几乎无法理解它的决策过程，因此它就是不透明的，比如特定输入的数据为什么要被特定标准进行分类，为什么产生了特定的输出等等问题”。但令人担忧的是，如果作为行政相对人的用户不理解为什么系统已经采用这种方式而不是其他方式，监管机构很难核实智能系统的某个具体决定是否遵守了法律的要求，简言之，缺乏透明度的智能系统阻碍了监管效果的实现。此外，不同于其他复杂技术的是，智能系统既是国家监管的直接对象，本身也参与行政监管的过程并“行使”国家公权力。但一个前置问题是，我们能否将行政职权委托给人类无法理解的智能系统？当决策不透明的机器出现在法律体系中，而当决策的可追溯性是一个重要法价值的时候，机器决策的不透明性就是一个重要的问题。

智能系统“黑箱子”的现状与政治学强调的最大程度的“透明度”相冲突。这要求受决策影响的人群应当深入了解系统及其决策过程。倘若无法保证系统决策的透明度，则应停止使用该系统，因为法治国原则和民主原则均是决策透明的理论基础。

当然，需要注意的是，对技术透明度的要求一定程度上是一种过于一般化的讨论，且这个命题本身也是十分值得怀疑的：智能系统是否一定比法律所代表的的传统行为控制模式更加不透明？在系统理论的语言中，这些行为控制模式是精神与社会系统。这些系统与智能（技术层面的）系统一样，具有内部动态性、自我指涉性和涌现性的特征，因此对于自己和他人而言就是不透明的。人，或是层级制度的治理模式、市场、谈判或网络控制着特定的行为，无论是直接控制实体还是间接控制实体，都会使得黑箱外的人无法准确预估决策的具体过程及其结果。在这种背景下，尼古拉斯·

卢曼将现代社会称为“不透明的交响曲”。即使在法理学中，长期以来也普遍存在这样的论断：控制不是机械的，而是一个复杂的、具有潜在不确定性的过程，也就是一个不透明的过程。

这一论断与智能系统决策透明度的要求并不矛盾。因为同样缺乏透明度的传统决策方式并没有像智能决策方式那样给相对人带来普遍的难以接受感。事实上我们无法否认，我们对机器提出了比对我们人类自己更高的道德要求。法律体系本身作为“不透明交响曲”的社会系统的一部分，在面对法治国原则和民主原则的时候也存在透明度的问题。但我们依然对智能系统的监管和透明度问题保持乐观态度，原因在于可以找到法律制度上的答案。

智能系统具备的一些特质也有助于实现前述规制目标：比如智能系统的可塑性。智能系统的主要特征并不是更高的不透明度，而是它可以通过程序比人类决策者或传统的行为控制模式更容易和更精确地对行为进行规划和安排。在此背景下，关于透明度的讨论可以更加深入。智能系统，作为整合社会与法律互动关系的力量，其决策逻辑一开始就是基于人们无法理解的社会和法律背景。在此之上讨论的决策程序透明度问题并不一定只限于纠正技术缺陷。反观透明度本身的内在要求，尝试优化定制系统也是有其特定意义的，不仅可以弥补现在机器决策的不足，还可以在总体上寻求更高层次的透明度。

前述的讨论是以一种长期的眼光看待智能系统的透明度问题，因为目前没有技术或法律要求以这种方式解决智能系统的透明度问题，真正能够回应的应对措施也不过是在法律层面正视、承认透明度问题，并以正确的方式看待它。第一步，有必要澄清谈论智能系统缺乏透明度的真正含义（下述第一点）。然后必须要探求是否真的不存在规制透明度问题的法律框架：通常学界认为《条例》规定了自动决策的解释权。但即使《条例》包含这样的规定（虽然人们很少谈及这一点），《条例》主观权利的方法并不能完全满足透明原则的要求，即给予涉事主体查看智能系统的代码的权利并不能解决任何问题（下述第二点）。在这一点上，了解宪法和行政法的传统教义学理论是有帮助意义的。因为在宪法和行政法的领域中，一直都存在事关处理决策过程的成熟方法，这也是 Burrells 所谓的“不透明”问题在人类决策领域的制度答案。为了加强“透明度”，法律体系不依赖于孤立的权利去解释决策的所有因素，而是从一个规则、制度的网络运作的视角切入。这些规则和制度的重点在于探寻理由上（下述第三点）。这表明一种新的尝试，即理由概念转换为监管手段并嵌入智能系统的决策过程。下面我们将概述控制架构中理由的主要特征，并论述如何将其转换进入智能系统中（下述第四点）。

（一）透明化——智能系统应用环节中的挑战

缺乏透明度从直观上看，是因为只有一小部分人可以阅读计算机代码，也只有专家团体才具备能力直接对原始系统进行分析。然而，由于 IT 部门组织分工精细化程度较高，即使是训练有素的计算机专家也无法事无巨细地了解整个计算机程序。对于基于大量数据做出决策的程序，系统复杂性会成倍增加。即使是具备确定性的决策系统，在决策过程中判断相关性都会处理海量的数据，人们在了解系统决策过程时也会被这些数据迅速淹没。

智能系统的复杂性在应用中也会不断地上升。因为在学习过程中系统会不断调整其变量的内部权重，使其在学习过程中对所做的决策进行反馈，并在之后的使用过程中反馈出来。系统在时间点 t_0 作出的决策结果，可能与在时间点 t_1 作出的决策结果完全不同。这种内部控制系统的动态变化，意味着对决策的任何解释仅在某个具体的时间点才有效。因此，拥有高质量即时响应能力的系统必然具有内部的高复杂性。

在这种情况下，“透明度”“可解释性”或“可理解性”等术语的含义是什么？这里可以先区分

不同的情况：（1）对于系统基本结构的解释，普通人已经难以轻易理解。（2）另一方面，具备专业知识且能构建相应系统的专家，能够通过审查系统逻辑——也就是系统的算法架构和数据库的结构使用的逻辑——对系统的算法架构和数据库结构进行根本性的理解和描述。（3）专家们难以实现具体决策。由于事先不知道“系统—环境”的相互作用，因此专家们对于动态智能系统，难以在事前进行决策预测。甚至在事后重建活动中也难以给出明确的“if-then”（如果……就……）关系，即很难查明特定因素对于特定输出的因果关联性。更确切地说，“决策规则”是基于动态数据集和大量加权变量概率函数的运算结果，其结果有时会存在进一步的数学过程，以获得最佳的计算时间。虽然可以记录这些变化和运行过程，但人类观察者还是会发现：机器学习运用数学模型优化算法的高维度特征方法，与人类推理的语义解释风格的并不能很好地搭配在一起。（4）当然，这种差异也使得我们更容易衡量使用智能系统的效果。如果记录了决策本身和它的效果，则可以在事后对数据库进行评估，例如特定的敏感群体是否承受了不成比例的负担，这会帮助我们识别哪些算法可能会存在歧视风险。

另外一点是——虽然这些过程不是技术层面的，而是合法性层面的——智能系统的透明与否。例如，许多智能系统的设计都要对有可能威胁到到企业商业秘密的人保密。德国联邦最高法院在对评分系统作出的决定进行裁判的时候基本确立了这一立场。欧盟《条例序言》的第63条第5项也是持此立场。除了保护知识产权之外，保密性的价值还在于应当防止他人对于系统的操纵。例如，《德国税法通则》第88条第5款第4句明确规定，财务风险管理系统的细节不得“在可能危及税收的规律性和合法性的情况下”被公布。

（二）通过主观的“解释权”能否达到透明化？

对于智能系统决策的相对人，智能系统的决策的不透明性体现得十分明显。当他们受到特定决定结果影响的时候，比如贷款被拒绝，税务审计遭到不利评价，或是被下令采取保全措施的时候，他们将面临几乎无法克服的技术与法律的壁垒。这种情况会给人们这样的一种印象：个人成为机器决策的对象。

数据保护法第一次回应了这个问题，《条例》第5条第1款规定了透明原则（前面已有论及）。《条例序言》的第39条、第58条和《条例》的第60条也都提供了实证法上的依据。《德国联邦数据保护法》《欧盟数据保护指令》，以及《条例》和《针对警察和刑事司法的数据保护指令》，都已经对自动化决策系统有了一般性要求，这些要求，也都是与透明度密切相关的。例如，《条例》旨在使数据主体能够审查其个人数据被处理时的合法性，以便于个人在必要的时候行使进一步的权利（见《条例序言》第63条）。鉴于最近关于算法和智能系统的争论，这个包含复杂结构的规范目前受到了极大的关注。特别是《条例》有时被解释为对机器决策在解释权意义上进行了极端的扩张。因此，透明度问题看起来已经找到了合法性依据。

1. 欧盟《条例》是否存在“解释权”？

首先，《条例》中并没有提到“解释权（Recht auf Erklärung）”。但是可以通过第13条到第15条，以及第22条对自动决策系统的要求推导出来。同时，《条例》在第13条第2款f段和第14条第2款g段都规定了控制者对于非学习型系统的用户承担信息提供义务，即需要向用户提供系统设计。这包括第22条第1款和第4款所提到的逻辑程序，以及对于数据主体的处理过程的重要意义和设想结果。第15条第1款h段授予了数据主体相应的信息权。第22条规定了算法决策系统的传播，在第3款中，规定了如涉及第2款a段和c段的情况，应当严格地限制自动的数据处理，应当加以合理的保护措施以及必要的程序控制——有人认为这应当包括解释权。

但是，应当谨慎对待此类观点。虽然上述提到的规范与《欧盟数据保护指令》（DSRL）第 12 条 a 段和第 15 条在语言上很相近，但是二者仍然存在本质差异。与老版本的《德国联邦数据保护法》第 6 条 a 款（新版本《德国联邦数据保护法》第 37 条）以及相同领域的平行规范——《德国联邦数据保护法》第 28 条 b 款关于特殊领域的评分系统（新版《德国联邦数据保护法》第 31 条）的规定一样，《欧盟数据保护指令》在学术和实践领域都被非常谨慎、保守地适用，对法律没有涉及的方面，保守的解释就已经足够。“信息（Informationen）”一般被限定为对于一般决策系统的简短评论，除非是该逻辑决策树被要求必须披露。另一方面，相对人对于原始数据、程序代码或算法很难深入了解，个案中都很难了解具体细节。因此，在涉及评分系统的情形下，有关人员仅能获得关于系统运作的初步信息，比如积累的分数与获得贷款可能性之间的关系。

《条例》下的判例法并不能改变上述的保守观点。第一，第 22 条的适用范围仍然很有限。尽管该条文适用于学习型系统和非学习型系统，但如果该系统仅仅用于辅助人类决策者，则第 22 条仍然得不到适用。第二，就条文的内容而言，按照通说并比较旧的立法例，这其中没有太多的漏洞（需要解释补充的地方）。例如，立法过程中讨论的“解释权”并未强制性的规定在第 22 条第 3 款，而仅仅是被作为可能的保护机制示例在《条例序言》中被提及。

《条例》的第 13 到第 15 条也是基于《欧盟数据保护指令》的第 15 条 2a 款。根据《条例》第 13、第 14 条，数据收集人的信息披露义务变得更加严苛，负责人必须在作出收集数据的决定前，披露相应的信息。特别是对于在使用过程中发生动态变化的、具备学习能力的系统。事前解释当然并不足以说明在个案情况下作出决定的原因。对于（后续）根据《条例》第 15 条规定的数据权利主体可以要求数据收集者提供数据的权利，则是与上述无关的另一套规则。然而，从规则发展史和体系关系的角度来观察，非常清楚的是，《条例》第 15 条赋予受到数据收集行为的主体更多的信息权利，而义务人必须根据第 13、第 14 条事先提供信息。此外，目前对于《欧盟数据保护指令》的解释，以及国家政策对于解释权的限制都会促进《条例》中对于条文解释的狭义化：特别是在保护智能系统制造商的知识产权方面，《条例》明确承认，要保护商业秘密和所有有关主体的个人数据。

因此，《欧盟数据保护指令》和《条例》的规定是有所不同的。总的来说，《条例》认为受智能系统决策影响的主体信息义务和信息权限仅仅限于关于决策系统运作的一般系列，但无权在个案中要求做进一步的解释。

2.“解释权”的有效性

鉴于欧洲法院在数据保护法方面一直以来的开放态度——比如在文本、体系和历史解释方面都欠缺支持，但仍然做出了“被遗忘权”的判决——不能排除欧洲法院仍然会在司法实务行使解释权，授予受智能系统决策影响的主体要求数据收集主体提供数据库、算法等详细信息的主观权利。但是从单个主体的角度来说，这真的会保护这些用户吗？

透明度问题，如数据保护规范所表明的那样，仅从智能系统用户与有关用户之间关系的角度来设计规范，那么，这种规范设计实际上是一种替代性方案，迫使前者为后者提供对于系统的“洞察力”。如此这般的制度安排虽有其独到的功能，但前述关于智能系统缺乏透明度的技术和法律原因的分析已经表明，受智能系统决策影响的人很难从中受益。相反的是：越是强调这种做法，越是使得公民难以从中受益，同时也难以使人们获得“对于纯源代码的洞察力”。大部分公民很难理解这种制度的实际价值。更可行的另一种方法是：为受影响的人提供关于系统的一般描述以及在使用过程中提供具有统计显著性的简化因子的列表。但这种做法也有问题，除非出现明显的错误或混淆，否则受影响的外行人无法对这些大量而琐碎的信息做准确的理解。基础的信息难以揭示智能系统根本的质量缺陷、隐藏的歧视和数据保护违规等问题，也难以帮助受涉主体对于智能系统用户（比如

数据收集企业）提起成功率较高的诉讼。

在不破坏监管有效性的情况下，也不能随意减少该技术的复杂性。为了成功识别申诉并进一步提出更正的建议，首先必须要收集有关系统和决策过程的详细信息。系统、数据库、算法和培训模型也必须在可被访问的状态。最终，调查的过程变得非常复杂，以至于大多数真正受到决策影响的利害关系人不太可能理解整个过程。总而言之，在许多智能系统中，似乎不可能对具体的决定提供一种技术性的且符合《条例》第 12 条要求的解释结果，即精确透明、简洁易懂、易于获取的解释方案。

因此，全面的解释可以为受决策影响的利害关系人提供对智能系统的理解并能使之有效地行使权利之观点不过是一个幻想。这种方法不仅低估了技术的复杂性，也高估了个人的专业知识水平。也有观点认为，在评分系统的运作过程中提供简化的信息至少会增加人们对技术的普遍信任。这样的观点同样不可靠，并且实际效果可能恰恰相反。因为在任何情况下，专业能力都不是通过“解释权”来传达的。社会问题是通过主观化而被外化的，而系统真正的运行方式则始终存在于黑暗之中。最后，“解释权”并未提供针对企业和商业秘密等复杂问题的解决方案。但上述这些并不意味着这种解释权对提高透明度毫无用处。但是，仅凭解释权无法解决信息不对称的问题。

（三）透明原则通过提供理由，实现国家的保障义务

上述是否意味着法律应当向技术举出小白旗吗？这个结论还为时过早。可预见的是，将透明度的问题私法化解决是不够的，将透明度问题的解决完全建立在系统用户和受决定影响的利害关系人的关系上，也是不够的。如果现有的方案无法达成目标，是否意味着透明度问题需要被重新定义或改变？

在这一点上，至少从受影响人的角度来看，后退一步有助于意识到现有制度存在的问题。如上所述，即使是受到系统决策具体影响的利害关系人，通常也很难理解自己作出的决定。每个人对于其他人都是“黑箱子”，这种情况存在于每个个体决策者行动之中。比如，人类通常对于刑事诉讼过程中的侦查犯罪并进行司法活动的行为都持有怀疑，刑事诉讼程序总被描述为易于出错和不可确定，且总被形容为“从根本上就不透明”。这与复杂的机器类似，刑事诉讼过程中人类官员的决策过程和具体细节，根据现有的制度框架，也是不可能被决定相对方真正透彻地了解的。

尽管上述法律制度均缺乏透明度，在无法用现有制度对所有关乎到决策的因素进行规制的情况下，致力于促进不同主体相互理解的宪政国家也并未处于崩溃的状态中。这表明，从法律的角度来看，对透明度的要求并不旨在对个人和集体决策过程进行全面的解释，也不在于对决策系统内部动态的检视。当人们将宪法和行政法作为法律传统的起点开始观察的时候，就会发现这样的一种观点是具备解释力的，即透明度旨在确保决策能够为决策对象理解，进而确保有关主体能够适当地维护其权利，简言之提供了“理由（die Begründung）”。

要求陈述理由是宪法国家的核心要求：当政府作出决定的时候，出于民主、法治和基本权利保护的原因，法律经常要求政府提供理由。《联邦行政程序法》强调了理由和监管行为之间的联系，并制定了“行政相对人有权知道行政行为的论证过程，因为这样他才能妥善捍卫自己的权利”这样的规范。类似的规范在欧盟法律和许多宪法性规范中都存在。

那么到底什么构成了理由？要采用哪种形式的解释，以便在法律层面确定某个决定的可追溯性？一旦这些问题得到正确解答，就可以将之用于解决机器系统的不透明性并促成相应的制度建构。

理由是指确定基本事实的活动，包括对证据的评估以及法律推理。在陈述理由的过程中，决策者需要解释“为什么他决定是这样而不是另外的情况，决定者可以运用经典的文献、教义、偏见或是一般的正义观念，或是运用论题学的方法，来进行法律解释”。关于理由的结构和内容的初步结

论，可以从上述这些基本的命题中得出，进一步的细节则可来自对文献、判例法中承认的法律推理和对制度功能的分析。

因此，对于陈述理由的要求首先就是行政主体进行自我监管的一种方式。行政主体首先被要求必须要陈述“基本的事实和法律原因（《联邦行政程序法》第 39 条第 1 款第 2 段）”来支持自己作出的决定，同时，这也会一定程度上审查这些理由是否包含在法律对该行政机关的授权范围之内。如果在陈述理由过程中，只是陈述不言而喻的事实，这将无助于自我监管。重要的是，需要明确在一个法律判断过程中会受到哪些因素的影响。即使是在裁量过程中，也必须能够追溯并明确哪些要素真正影响该决策。这样一来，如果公权力再次做出决定，就必须符合他们之前所提出的理由，符合他们自己的标准。从自我监管的角度来看，解释性的理由并不在于对完整而客观的理由进行厘清，而是展示决策之中的关键性要素。这也需要公共机关的若干程序性义务加以配合，比如作出决定的机构应当记录自己的论证路径与过程。

此外，提供理由还有助于对决策的外部监管。它的价值在于给予决定相对人机会，来了解决定的合法性。当他们寻求法律保护的时候，公权力作出的论证理由将会指导他们针对特定的要点进行申诉。然而，最重要的是，通过提出理由，可以审查决定作出的过程，进而监督行政机关和法院。理想情况下，在论证过程中，是允许决定的机构独立地对事实和法律问题进行补充和纠正的。之后，通过法律职业团体的工作，理由得以进入法教义学的话语中，从而超越了具体的法律决策场景。因此，理由是法律体系内的工具，它旨在对公权力机构的不同命令进行程序性监督，并促进不同实体之间的决策反馈，最终提高程序正义和决策质量。当理由的潜在接受者不仅包括受决定影响的利害关系人，还包括监管机构和法律职业团体时，对于复杂问题的决定的法律论证，也必须对问题的复杂性自我证成，这样才能进行有效的监督。因此，理由的特点和程度取决于个案的情况，而不是有关人员的个人能力。如果作为外行人无法理解法律论证，则似乎应当寻求法律建议。但事实恰恰相反，在更高级别的权力机关作出的决定和理由可能会具有普遍性，这时则需要排除个案的情况论证理由，以使得法律问题与一般的政治、民主话语保持一致。

理由是社会接受和系统信任的产生机制。理由在一定程度上有着促进商谈的功能。实际上，理由对于社会接受和信任能产生多大程度的积极影响，是一个经验性问题。虽然没有确切的数据，但似乎也能得出以下比较合理的结论：政府越是努力地解释做出决定的理由，并且它在听证程序中越多地回应利害关系人所提出的论点，决定相对人也就越容易接受特定决定行为。从长远来看，这也有利于社会对于该机构的普遍信任的建立。在智能系统的决策中更能体现这一点，完全拒绝向相对人提出理由会使人们对智能系统的决定产生阻力，并危及人们对技术的信任。对于行政透明度的要求会有助于这种情况的解决，并建立普遍信任。为了形成惯例，必须要求行政主体能让相对人感受到具体的说理和论证——即便只是简单的理由，也会传递出行政主体的可沟通性，以及该行政决定的可救济性。

理由同时也是对决定相对人进入法律程序、表达利益诉求的一种承认。即便作为公权力机关的自我监管的工具时，论证有时会失能，即便决策的相对人在没有专业支持的情况下，既无法理解理由的内容也不接受更为具体的理由，但只要理由制度存在，国家与公民的关系的核心仍然是基于确信（Überzeugung），而不是建立在威压之上；正如尼古拉斯·卢曼所言，理由总是单独地被确信的建立者记录下来。因此，理由是一个命令的核心要素，它旨在促进论证和反对论证的相互交换。行政决定的理由往往是贯彻法治国原则的第一要素。

在不同的教义学理由的帮助下，立法者、司法裁判和法学教育保证了法律理由在不同情形下的正常运行。根据主体、时间、领域和行为的不同，理由也会不同，这也进一步区分了不同的标准，

比如不同的法律地位、有关的事实和不同的行为模式等要素。在此时只有两点是最重要的。

第一，法律论证中的理由并不是要解释所有影响决策结果的要素。因此，在法律的语境下，社会学、心理学、生物学、物理学等因果要素是无关紧要的，虽然从后科学社会的视角来看，这些要素都可以帮助解释一个决定。但这并不意味，这些因素在法律中不起作用。政府官员和法官在决策者的情境中不仅仅受到法律理由的指导，还受到众多偶然因素的影响。而这些偶然因素也会使官员受到相当大认知上的扭曲，现实主义法学就基于此对法律论证进行了批判，其观点也在晚近的文献中得到了经验上的验证。尽管有这些经验研究成果，但法律论证至今为止仍保持了其内在逻辑的自足和证立。因此，理由制度仍在法律体系中具有独特的地位。

第二，决策的理由与决策的解释是不同的。因此，对于一个决策所附带的论证并不当然等同于对一个决策合法性的客观判断。在有疑问时，理由并不终局性地解决争议问题，而是启动了法治国家的程序机制去解决这个问题。据此，理由的制度性要求是基于对法律和法治国的一种理解，即对法律是针对社会分工的一种监管措施，通过设定一种理由的文化——促进理由和反对理由的意见交换——建立并强化人们对于行政命令的基本信任。为了保护基本权利、维持法治国的稳定并实现合法化的效果，理由设置了一种差异化的制度体系，在这个体系中，法律意义上正确的决定是建立在不同机构间分权的基础上的。在这个意义上，对于陈述理由的要求不仅仅有利于有关人员，也有利于整个系统。每一个理由，都在守护民主原则要求下的透明性原则和法治国原则要求下的合理性原则。

当然，对于陈述理由的要求也并不是绝对的。在德国法律体系中就存在很多例外。最为重要的一个例外是在议会法中，但是它也依赖于特定的合法性来源，并受到自身监督机制的管控。最具争议的问题是行政指令是否有提供理由的义务。值得注意的一点是，大多数人能够接受行政和司法决定领域的众多例外。当然，这都是基于一些特别的情况，例如决定的结果可能对当事人的影响微不足道（《联邦行政程序法》第 39 条第 2 款第 2 段）或是受决定影响的相对方的信任（《联邦宪法法院法》第 93 条第 1 款第 3 段）。

（四）通过分工机制的建立和控制架构的结合，达到智能系统的透明化

针对人类决策发展出的规则和标准，不能强行地应用于机器决策。人们普遍认为，根据现有的制度情况，执行宪法义务使得智能系统的每一个决定都能被相对人所理解是非常困难的。因而有必要将智能系统整合到“提出理由和反对理由”和“决定—监管”的系统中，即融合到现行立法体系中。只有实现上述情景，才能促进人们对于技术长期的信心增长，也只有这样才能实现技术的潜在生产力。

上述这种方案对应的技术目前还未发展成熟，且并不存在将智能系统纳入法律制度的蓝图和最佳方式，但论证可以提供一个基础，在此之上可以从法律的角度有效地批判机器决策。

这一方面确保了智能系统在与人交互的过程中可以自我监管自身的决策过程。目前很多项目都在尝试开发技术解决方案，为智能系统提供自我监管机制和自省机制。这些项目的长期目标是：以一种特定方式对智能系统进行编程，使得它可以自己提供有关支持特定决策的理由，并以一种可让相对人理解的方式呈现信息。对于开发这种所谓的“具有解释能力的人工智能（Explainable AI）”到底能否成功，结果还无法预料。因此，我们不单单要关心程序本身，也要重视编程环节。目前正在讨论的各种智能系统设计者的行为准则，一定程度上也是在完善智能系统的自我监督。虽然上述的监督（Audits）是在企业的倡议下进行的，但实质上仍然属于智能系统自我监督这一领域。此类监督或审核是一个在不断发展的产业，也是一个新兴的研究领域。这种审核实质上是通过评估数据库结果集或使用“虚拟数据（dummy data）”和控制算法来检查智能系统。前述这些技术都是在业

界运用了很长时间的软件控制技术。虽然所谓的“黑箱测试”无法解释每一个具体的决定，但是通过它可以获取关于系统效果和一般决策行为的经验，并诊断质量缺陷或隐藏的歧视，这也同时可以对系统错误进行纠正。立法者可以用一些举措来达到这样的效果，例如定期进行审核（Audits）或采取其他的措施，比如公布源代码，或对企业进行减免责任以资奖励。

仅靠自我监管是不够的。因此，国家机构也可以对智能系统的决策进行外部控制，同时，有必要在法院和行政系统人员内部适当普及智能系统知识。从其他领域的经验来看，对于该领域的规制是一项长期艰巨的任务。智能系统的规制目标的实现路径自己就是政府应当参与到公私专家网络（öffentlich-privaten Experten-Netzwerken）和技术标准化机构的工作中去，通过以上这两种方式将智能系统的知识传递给国家机器。下一步则是建立专门的机构，在能获得社会专业知识的情况下，自己也能获取相关的知识。随着专业能力的提高，这些机构本身也可以制定技术标准，比如内部一些行政标准，并在政府和私人的背景下对智能系统进行认证和审核（Auditierung）。可以想象的是，这种预防监管制度可以按照不同方式构建：一种是用责任特权来奖励成功的认证；另一种则是根据不同的应用领域和软件产品投放领域来设置不同的行为要素，比如企业必须提供相应的证明，来证明自己的系统做到了无歧视，而证明无歧视的过程中，必须结合自己系统的应用领域和产品投放领域进行说明。另外还可以通过公权力推动的审核（Audits）专门用于事后对算法进行审查，整个程序的启动既可以依照职权启动，也可以依照行政相对人的申请启动。监管过程中的协助义务和信息披露义务促进了监管效率的提高。虽然以上所有提到的措施不能完全防止错误和权力滥用——就像现有的宪法监督制度不能完全消除人类决策者的错误和权力滥用一样，但是以上的这些程序可以降低错误和权力滥用发生的概率，并限制损害的范围。

对于外部监管而言，有一个非常重要的方面，就是如何刺激相对人启动监管程序。要考虑到相对人之间在知识背景等方面的不同，因此解决这一问题可以采用渐进分类的透明制度：对于直接受到决定影响的相对人，应当告知其系统的一般功能；个案情况下可告知与之特别相关的一些因素。但是，如果案涉相对人在此基础上要求进一步审查该决定，则必须确保他们或其指定的专家在程序进行期间能够获得所有的必要信息。这可能包括根据需要披露的数据库、算法和机器训练程序。决策系统的社会功能越重要，这种分析就必须越深入和透彻。从公权力机关和法院的角度来看，至关重要的是，这些程序原则上应当遵守程序法或行政法要求的既定形式和程序，以便随着事件的深入，那些不具备专业知识的公权力机关（比如监管部门和司法机构）可以进入程序并制定监管的标准。

最后，关于透明度方面，要想让新技术被社会普遍的接受，政府、社会和个人要对超越法律话语的新技术的发展可能性和风险进行社会和科学的审查。类似像“算法监控者（AlgorithmWatch）”这样的社会活动已经在这一方面发挥了重要的引领作用。媒体也应监督和批评智能系统的应用和发展。最后，国家肩负着公共教育的使命，也应当将适当的内容传播给公民。

但是，人们也不应当忽视，每一项对透明度的制度要求都有其局限性。智能系统与人类决策在这一方面并没有什么不同。然而，“目前的法学理论在智能系统参与行政决策面前会遇到新的挑战”这一命题是值得怀疑的。判例法和立法者迄今为止对智能系统用户的操作行为和商业秘密，都表现出了相当的重视和灵敏度。在此背景下，他们提出监管和限制智能系统的深层次理由，例如知识产权保护、数据保护、安全和其他公共利益。作为一般性规则，这些原则都必须服从于透明度原则；越是要求灵敏度的领域，也越有可能应用智能系统。但目前的做法过于僵化，没有必要在所有的情境中要求智能系统对自己的决定进行解释，也没有必要在所有情况下都对智能系统的决策进行监督并要求相关方提出理由；这是没有效果的。也就是说，在智能系统这个领域要求绝对的透明度和理由是不可行的。无论如何，监管智能系统必须要有平衡透明度和保密利益的措施。可以考虑引入

“相机程序（camera-Verfahren）”或是在保密义务下的专家监督员制度。

权属所有措施的成功与否都取决于法律的政策倾向性，虽然这些措施会损害个别案例中的创新企业。在这些议题的讨论中，经常可以看出私人利益对于公共政策有着非常大的限制，诸如Alphabet、Facebook和亚马逊等常被讨论的头部企业现在恰恰正是智能系统的最大投资者之一。然而，私人权力（private Macht）是否真的是政府监管的障碍，或是否导致了对于少数、突出、影响巨大且易于识别的主体监管失能，答案并不是绝对的。这些问题的答案，很大程度取决于跨国公司能否充分利用全球法律的差异化，但最终的结果均是人们对私人权力会阻碍国家行使职责的担忧。

前景与展望

本文概述了在智能系统中应当嵌入的理想化监管规范架构，以此确保智能系统的安全性，进而促进人们对于智能系统决策的可接受性和信任程度。并且，监管规则应当具备强适应性和灵活性，依据系统使用环境、使用主体的公私属性提供不同的救济方案。这种监管模式不至于成为技术创新的新障碍，相反，还能够为智能系统技术开发指明朝向社会福祉的正确方向。此外，有必要建立一个根据决策理由进行监督、审查和评判的整体性监管架构。如果在未来人们成功地将机器决策纳入到宪法制度、法律程序，以及提出理由、反对理由的意见交换制度框架内，那么，个人在智能系统面前就不再是弱小的客体，而可以成为真正的主体。这种参与机制会促进个人摆脱对技术的客体式依赖，通过特定主体对技术和决策的具体评价内容真正确保个人成为数字化秩序的主体。

Abstract: The large-scale application of artificial intelligence (AI) system poses a great challenge to the law. At present, the development of intelligent system has not reached the level of human intelligence, the unpredictability and uncontrollability of AI system will cause a series of problems, so it is necessary to supervise the application mode of AI system. At present, there are three core technical elements of AI system, namely, data, algorithm and human intervention. These three elements play an anchoring role in the establishment of regulatory principles, and they are also the focus of the implementation of the regulatory system and the regulatory object of legislative adjustment. The biggest obstacle to supervising the AI system is that its decision-making process is difficult to be understood by the opposite person, which means that AI system needs to maintain a high degree of transparency in the decision-making process to ensure understandability. However, high transparency is difficult to achieve through ex post decision interpretation, but should be achieved through the decision-making system to provide decision reasons. Therefore, it is necessary to establish a reason and regulatory framework based on intelligent system decision-making, in which machine decision-making is incorporated into the legal process. Such a framework can make the decision-maker and the other party fully exchange views, so that the individual can really become the subject of the digital order, rather than the object.

Key Words: AI; Intelligence System; Transparency; Legal Supervision

（责任编辑：李广德）