

南开大学

科技法学结课论文

中文题目：算法歧视的反歧视法体系规制分析

外文题目：Analysis of the Anti Discrimination Law
System for Algorithmic Discrimination

学号：2113203

姓名：付政烨

年级：2021 级

专业：信息安全、法学双学位

系别：信息安全系

学院：网络空间安全学院

指导教师：刘芳

完成日期：2024 年 6 月

摘要

本文探讨了在人工智能（AI）和算法决策日益普及的背景下，反歧视法律体系如何有效规制算法歧视问题。首先，介绍了算法及人工智能的内涵，并区分了算法决策中的歧视与差异化。针对算法可能导致的两类问题：对特定人群的伤害性差异化和非特定人群的不公平差异化，本文深入分析了当前反歧视法律体系面临的挑战。通过比较完全开放、完全封闭及混合法律体系的特点和适用性，本文指出混合法律体系在应对算法歧视中的优势，尤其是在结合法规编码技术时的潜在效用。混合法律体系不仅能提供明确的法律指导和确定性，还具备灵活应对新型歧视形式的能力。结合法规编码技术，混合法律体系能实现自动化合规性检查，确保算法在设计和应用中符合法律要求。本文最后总结了未来研究的方向，包括加强法规编码技术的开发、推动跨学科合作以及促进国际合作，以构建一个公正、透明且包容的人工智能反歧视法律框架，为社会的可持续发展提供保障。

关键词：算法歧视；反歧视法律；混合法律体系；法规编码技术

Abstract

This article explores how anti-discrimination legal systems can effectively regulate algorithmic discrimination in the context of the increasing prevalence of artificial intelligence (AI) and algorithmic decision-making. Firstly, it introduces the concepts of algorithms and AI, and distinguishes between discrimination and differentiation in algorithmic decision-making. The article delves into the challenges faced by the current anti-discrimination legal system in addressing two types of issues that algorithms might cause: harmful differentiation against specific groups and unfair differentiation against non-specific groups. By comparing the characteristics and applicability of fully open, fully closed, and hybrid legal systems, the article highlights the advantages of hybrid legal systems in addressing algorithmic discrimination, especially in conjunction with regulatory coding technology. Hybrid legal systems not only provide clear legal guidance and certainty but also possess the flexibility to respond to new forms of discrimination. Combined with regulatory coding technology, hybrid legal systems can achieve automated compliance checks, ensuring that algorithms comply with legal requirements in their design and application. The article concludes by summarizing future research directions, including enhancing the development of regulatory coding technology, promoting interdisciplinary collaboration, and fostering international cooperation to build a fair, transparent, and inclusive AI anti-discrimination legal framework, thereby ensuring sustainable social development.

Key Words: Algorithmic Discrimination; Anti-Discrimination Law; Mixed Legal Systems; Legal Coding Technology

目 录

摘要.....	I
Abstract.....	II
目录.....	III
一、 引言.....	1
二、 算法歧视和差异化.....	2
(一) 算法、人工智能的内涵.....	2
1. 算法	2
2. 人工智能	2
(二) 算法歧视和差异化的内涵.....	3
1. 伤害特定群体的差异化	4
2. 不伤害特定群体但仍不公平的差异化	5
3. 差异化造成的危害后果分析	5
三、 反歧视法律体系.....	7
(一) 反歧视法律体系的内涵.....	7
1. 完全开放体系	7
2. 完全封闭体系	7
3. 混合体系	8
(二) 中国与欧盟的现有法律架构梳理.....	9
四、 遏制人工智能相关的反歧视法律体系.....	11
(一) 完全封闭、开放的法律体系难以适用算法歧视.....	11
(二) 应对算法歧视的最佳反歧视法律体系.....	12
(三) 混合法律体系利于与法规编码技术相结合.....	13
1. 编码法规技术	13
2. 混合的反歧视法律体系与法规编码技术结合	14
五、 总结与展望.....	16
参考文献.....	17
致谢.....	19

一、 引言

算法决策和人工智能的广泛应用，确实在社会的诸多领域带来了显著的改进。然而，伴随这些技术进步的同时，也不可避免地产生了一些歧视性影响。尽管现有的反歧视法律在某种程度上为公众提供了保护，但 AI 决策带来的新挑战正逐渐显现^[1]。例如，一些算法可能会根据用户的网络浏览器偏好或公寓号码等看似无关紧要的特征，对人群进行新的分类，或者通过多个数据点的组合形成更为复杂的类别。在荷兰，某些保险公司甚至会针对住在特定类型数字（例如带有字母的数字，如 5A 或 5C）公寓的客户，收取额外的汽车保险费。

在这样的背景下，本文探讨了以下几个重要问题：当了解到算法可以基于与受保护特征（如种族或性别）无关的特征进行区分时，什么样的反歧视法律体系最适合应用于算法决策？此外，本文还进一步探讨了立法者应如何应对基于非受保护性基础的算法差异化的最佳途径。在此处对本文涉及的一些常见概念进行解释，具体内容如下表所示：

	定义
歧视	对具有受保护特征的人造成伤害，构成不合理的平等待，或者被广泛认为是不可接受的行为
差异化	中性意义上的歧视
算法决策	基于计算机算法输出的决策

图 1.1

本文结构如下：第一部分介绍了算法决策及其相关概念，明确区分了算法决策中可能带来的两类风险：歧视和不公平差异化。第二部分详细探讨了反歧视法律的基本框架和其在不同地区的具体实施情况。第三部分分析了反歧视法律体系中受保护基础的封闭和开放列表之间的区别。第四部分则进一步探讨了这两种列表的特点、优势和劣势，揭示其在实际应用中的具体表现和效果。第五部分考察了在面对算法歧视时，哪种反歧视法律体系最为有效，并提出了混合法律体系结合法规编码技术的应对策略。最后，第六部分总结了研究成果，提出了未来研究和实践的方向。

^[1] Sweeney, L.: Discrimination in Online Ad Delivery, Communications of the ACM, 2013, Vol. 56, No. 5, pp. 44-54.

二、 算法歧视和差异化

(一) 算法、人工智能的内含

1. 算法

算法是指一个计算过程的系统化描述，通常表现为计算机程序，用于处理输入数据并产生输出结果^[1]。在计算机科学中，算法是解决问题或执行任务的一系列步骤，其目的是使计算机能够有效地执行特定的操作。一个算法应当具备以下几个关键特征：

1. **确定性 (Determinism)**：每一步操作都有明确的定义，即在算法的每一步骤中，计算机所执行的操作都是明确和唯一的，不会存在不确定性或模糊性。
2. **有穷性 (Finiteness)**：算法必须在有限的步骤内完成，即不会无限循环。
3. **输入 (Input)**：算法接收外部提供的输入数据，这些数据可以是零个或多个。每个输入代表不同的变量或数据集，通过这些输入，算法能够获取所需的信息进行处理。
4. **输出 (Output)**：输出是算法执行后的成果，它体现了算法处理数据的结果。输出的正确性和有效性是衡量算法质量的重要标准之一。
5. **可行性 (Feasibility)**：要求算法的每一步操作都能在有限的时间和资源内完成，不会出现无法执行的情况。

算法作为计算过程的系统化描述，其目的是通过一系列明确、有限和可行的步骤，实现对输入数据的处理并产生输出结果。算法在计算机科学和工程中具有广泛的应用，包括排序、搜索、图像处理、机器学习等多个领域。

2. 人工智能

人工智能 (Artificial Intelligence, 简称 AI) 是一个跨学科的科学工程领域，其目标是开发能够模拟、延伸甚至超越人类智能的计算机系统^[2]。人工智能系

^[1] Leese, M.: The new profiling: Algorithms, black boxes, and the failure of anti-discriminatory safeguards in the European Union, Security Dialogue, 2014, Vol. 45, No. 5, pp. 494-511.

^[2] Russell, S., & Norvig, P.: Artificial Intelligence: A Modern Approach, Prentice Hall, 2009.

统通常通过模仿人类的思维、学习、规划、语言处理、感知和动作等方面的能力来实现。这些系统旨在通过对环境的理解和与环境的互动，自动完成复杂任务，并在不断变化的条件下做出决策。

人工智能的发展可以追溯到 20 世纪 50 年代，当时著名计算机科学家艾伦·图灵（Alan Turing）提出了图灵测试，作为衡量机器是否具有智能的标准^[1]。1956 年，达特茅斯会议被认为是人工智能研究的正式起点。在接下来的几十年里，人工智能经历了几个发展阶段，包括初期的乐观预期、随后的“AI 寒冬”（由于期望过高与技术进展缓慢之间的落差），以及近年来的复苏和快速发展^[2]。

人工智能涵盖了多种不同的分支学说，包括符号主义、连接主义、行为主义和混合方法，每个分支都侧重于实现智能行为的不同方法和技术：符号主义基于符号表示和逻辑推理，连接主义基于人工神经网络和学习算法，行为主义关注智能体在环境中的行为和适应性，混合方法则结合多种方法的优势^[3]。现代主流的人工智能方法主要集中在机器学习、深度学习和强化学习，通过数据驱动、神经网络和环境互动等方式，实现高水平的自动化学习、模式识别和策略优化，广泛应用于各个领域。

当今主流的观点将人工智能定义为：一种研究如何使计算机系统展现出类似于人类智能行为的科学与工程领域。人工智能系统通过模拟人类的思维过程和行为，来执行复杂的任务和解决问题。AI 的核心在于开发能够感知环境、学习和推理、并采取行动的算法和模型。

（二） 算法歧视和差异化的内含

算法决策往往被视为理性、高效且无懈可击的象征。很多人或许会认为，算法本身是中立的。然而，遗憾的是，算法决策也可能带来基于种族或性别的歧视，或者引发其他形式的不公平差异。尽管算法背后充满了逻辑与精密计算，但其应用过程中却难免存在隐含的偏见与不公。我们必须承认，算法并非总能超越人类偏见，反而在某些情境下可能放大这些偏见，导致更为复杂的歧视和不平等。因此，在享受算法带来的便利和效率时，我们也需警惕其潜在的社会风险，

^[1] Turing, A. M.: Computing Machinery and Intelligence, Mind, 1950, Vol. 59, No. 236, pp. 433-460.

^[2] 宋乐、林德康：《人工智能：从图灵测试到深度学习》，科学出版社，2020 年印行。

^[3] 李开复：《人工智能：未来发展的机遇与挑战》，清华大学出版社，2019 年印行。

确保技术进步真正服务于所有人，而不是成为不公平的帮凶。

算法差异化给反歧视法律带来了几个挑战。我们重点关注两类问题：一类是伤害特定群体的差异化；另一类是不伤害特定群体但仍不公平的差异化。

1. 伤害特定群体的差异化

我们首先探讨第一个问题，即算法应用中的差异化对特定群体造成伤害的情况。当一个算法应用基于包含歧视或偏见的人类决策的数据集进行训练时，这种无意的算法歧视便可能发生。

例如，2013 年的一项研究表明，谷歌用于向搜索引擎用户投放广告的系统受到偏见人类行为的影响。当人们在谷歌搜索引擎中输入具有非裔美国人名字时，谷歌显示的广告暗示这些人有犯罪记录。而当搜索具有白人名字时，这类暗示犯罪记录的广告显著减少^[1]。这可能是由于谷歌的算法分析了点击率最高的广告，从而继承了种族偏见。

这是因为用户在互联网上的行为往往反映了社会中的偏见和刻板印象。例如，一些人可能会更频繁地点击涉及犯罪记录的广告，这种点击行为本身可能带有种族偏见。当这些行为数据被用来训练广告算法时，算法会无意中学习并强化这些偏见。其次，训练算法的数据集如果包含了历史上的歧视和偏见，这些偏见会被算法继承并反映在其决策中。例如，过去的广告展示和点击数据如果存在种族偏见，算法在优化过程中会不自觉地强化这些偏见。这一现象表明，虽然算法本身是中立的，但其依赖的训练数据和优化目标可能会导致偏见的继承和放大。

从法律角度来看，这种无意的歧视通常被归类为间接歧视。一般来说，间接歧视指的是一项看似中立的实践，但实际上对特定群体（如种族或性别）造成了不利影响。因此，现有的法律框架提供了一些保护措施，以应对这种间接歧视。例如，《欧洲人权公约》第 14 条和《欧盟基本权利宪章》第 21 条均明确规定，受害者可以通过提供证据，证明某项表面中立的政策或做法对其群体造成了特别不利的影响，从而要求纠正不公^[2]。

^[1] Sweeney, L.: Discrimination in Online Ad Delivery, Communications of the ACM, 2013, Vol. 56, No. 5, pp. 44-54.

^[2] Makkonen, T. (2007). European Handbook on Equality Data. *European Commission*.

2. 不伤害特定群体但仍不公平的差异化

在实际应用中，绝大多数算法通过基于非特定特征（如汽车型号、居住地邮政编码等）进行差异化，提取事物及其相互间的内在联系。这些特征与保护特征（如种族或性别）并无直接关联。通过分析海量数据，算法能够揭示这些特征之间的复杂且不明显的相关性，并利用这些相关性进行预测或区分群体和个体。

例如，一家在线购物平台的运营商发现，使用“京东”应用程序的用户倾向于购买价格较高的产品，而使用“淘宝”应用程序的用户倾向于购买价格较低的产品。基于这一发现，该公司决定对使用“京东”应用程序的用户收取更高的运费。尽管这一差异化并不是基于种族或性别等受保护特征，但它可能导致使用某一款应用程序的用户被要求支付更高的费用。这种差异化可能被认为是不公平的，因为其依据并非个体的实际购买能力或其他合理因素，而是基于一个看似无关紧要的特征，即用户所使用的手机应用程序^[1]。

这种现象背后的根源在于算法的运作方式。算法通过分析大量数据，试图找到能有效区分用户行为的特征。尽管这些特征本身并不具备任何歧视性，但在特定情境下，可能会带来不公平的结果。在上述例子中，算法可能会发现某些特定的购买行为与用户使用的特定应用程序有关，但这种关联实际上只是表面现象，背后可能隐藏着更深层次的社会经济因素。算法在做出决策时，未必能全面考量这些复杂的背景，从而导致决策结果具有潜在的不公平性。

需要指出的是，这里我们的目标并不是要说服读者每种形式的算法不平等对待都是不公平的；不公平性取决于特定情况的所有环境。只是说明，即使差异化不一定会伤害具有受保护特征的人，算法差异化仍可能导致不公平的做法，或者至少是有争议的。

3. 差异化造成的危害后果分析

算法歧视不仅可能导致结果与算法设计者的初衷背道而驰，还可能在无形中强化社会的不公平现象。设计者们本意是通过技术进步来促进社会公平，但实际上，算法的偏差却可能引发严重的不公正结果。算法的差异化不仅有可能加剧

^[1] Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

社会经济群体之间的既有不平等，还可能固化现有的贫富差距。尽管设计者的初衷是利用算法的力量消除偏见，提升社会的整体公平性，但在实际应用中，算法往往因为数据的偏差或设计的局限性，而产生了意想不到的歧视性影响。

比如，某在线教育平台利用算法进行个性化定价，希望通过经济补贴使得经济困难的学生获得更多优惠，而让富裕学生支付更高的费用。然而，这一算法根据学生观看网课的次数来调整优惠额度，结果却背离了设计初衷。现实中，城市中家庭条件较好的学生往往拥有更丰富的线下教育资源，因此他们对在线教育的依赖较少。而在偏远山区，缺乏优质线下教育资源的学生则更多地依赖在线教育，从而导致他们观看网课的次数更多。这样一来，算法反而让低收入群体支付了更高的教育费用，而高收入群体却享受了较低的价格。此种算法差异化不仅未能实现公平，反而加剧了贫富差距，使经济困难的家庭背负了更沉重的经济负担，进一步限制了他们获取教育资源的机会。

更为严重的是，这种算法引发的差异化可能在不知不觉中固化既有的社会结构，使得社会流动性下降，贫富差距扩大。原本通过技术革新所期望达到的普惠性目标，反而可能被数据和算法的无形之手扭曲，变成了新的不公平制造者。同时算法差异化的这种“穷人付出更多”的现象不仅有违社会公平正义，还可能对经济困难群体产生深远的负面影响。Eubanks 警告称，算法可能会对经济困难群体产生许多不利影响，进一步加深现有的贫富差距，并可能为处于社会弱势地位的群体带来新的挑战，影响他们的幸福感和人类尊严^[1]。

^[1] Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

三、 反歧视法律体系

(一) 反歧视法律体系的内含

反歧视法律体系旨在通过法律手段保护个人和群体免受基于种族、性别、宗教、民族、年龄、残疾、性取向等各类特征的歧视。其核心是平等原则，确保所有人在法律面前一律平等，不因个人特征而受到不公平待遇。其中歧视包括直接歧视和间接歧视。直接歧视是指基于某个受保护特征的明显差别对待，而间接歧视则是指看似中立的政策或做法对某些群体造成不利影响。

1. 完全开放体系

完全开放的歧视法律体系以其包容性和广泛适用性著称。在这种体系中，法律并未明确列举具体的受保护歧视理由，而是采用一种普遍的、开放性的反歧视原则。

欧盟宪章第 20 条就是这一体系的典型代表，该条款仅简单陈述“法律面前人人平等”，未具体列举任何受保护的歧视理由或豁免情况。这种开放性设计的主要优势在于其高度的灵活性，能够涵盖各种形式的歧视，包括那些未被明确列举或尚未出现的新型歧视形式，从而确保法律的适用范围广泛且适应性强^[1]。

然而，正是由于其广泛的开放性，这种体系也带来了法律适用的不确定性。缺乏具体的指导条款可能导致法律适用的一致，增加司法解释的复杂性，并可能引发滥用。此外，由于没有明确的标准和界限，法院和执法机构在具体案件中需承担更大的解释责任，可能会影响法律的稳定性和可预测性。

2. 完全封闭体系

完全封闭的歧视法律体系以其明确性和可预测性为特点。在这种体系中，法律详细列举了受保护的歧视理由，并且对可能的豁免情况进行了明确的规定。这种封闭性设计旨在通过提供具体的指导和标准，确保法律的适用具有高度的确

^[1] European Union Agency for Fundamental Rights. (2010). The Fundamental Rights Report. Publications Office of the European Union.

定性和可操作性。例如，欧盟的《就业平等指令^[1]》和《种族平等指令》分别详细列举了禁止在就业中基于宗教或信仰、残疾、年龄和性取向的歧视，以及禁止基于种族或民族出身的直接歧视^[2]。

完全封闭体系的主要优势在于其法律的清晰度和确定性。通过明确规定哪些类型的歧视是被禁止的，这种体系能够有效地指导相关方的行为，并减少因法律不明确而引发的争议。此外，由于有具体的条款和标准，法院在处理歧视案件时也有更明确的依据，从而提高了司法效率和一致性。

然而，这种高度确定性的体系也存在一定的局限性。由于其保护范围是预先设定且有限的，完全封闭的体系在面对新出现的歧视形式或未被预见的社会变化时，缺乏足够的灵活性和适应性。此外，过于严格和具体的规定可能无法及时反映社会对某些新兴歧视形式的认知和保护需求，从而导致法律保护的滞后性。

3. 混合体系

混合体系结合了完全开放和完全封闭体系的特点，试图在法律的灵活性和明确性之间找到平衡。这种体系通常列出部分受保护的歧视理由，同时通过特定措辞保持对新型和未列举歧视形式的开放性。以欧盟《基本权利宪章》第 21 条为例，该条款明确列举了性别、种族、语言等具体的歧视理由，但通过使用“任何其他意见”等开放性词汇，暗示该清单并非穷尽，从而保留了对未来可能出现的歧视形式进行保护的可能性^[3]。

混合体系的优势在于其较强的平衡性和适应性。首先，列举的具体理由提供了法律的明确性和指导性，便于相关方理解和遵守。其次，通过开放性措辞，该体系具备了一定的灵活性，能够及时应对和保护新出现的或未被预见的歧视形式。同时，这种体系在维护法律的稳定性和可预测性方面也表现良好，因为它为法院和执法机构提供了具体的指导标准，减少了解释的不确定性^[4]。

混合体系也存在一定的劣势。由于既包含封闭条款又包含开放条款，这种体

^[1] Barnard, C. (2012). *EU Employment Law*. Oxford University Press.

^[2] European Commission. (2000). Council Directive 2000/78/EC Establishing a General Framework for Equal Treatment in Employment and Occupation. *Official Journal of the European Communities*.

^[3] European Union Agency for Fundamental Rights. (2018). *The EU Charter of Fundamental Rights: A Commentary*. Hart Publishing.

^[4] Craig, P., & de Búrca, G. (2015). *EU Law: Text, Cases, and Materials*. Oxford University Press.

系可能增加法律适用和司法解释的复杂性。法院在具体案件中需要综合考虑列举的理由和开放性条款，进行更为细致的解释和判断，可能导致法律适用的一致性受到影响。此外，半开放的条款也可能导致在特定情况下存在模糊地带，增加法律操作的难度。

总体而言，混合体系在提供法律明确性和保护广泛性方面表现出色，适应性强，但需要在法律适用过程中谨慎处理其复杂性和潜在的不一致性，以确保其平衡性和有效性。

(二) 中国与欧盟的现有法律架构梳理

	中国		欧盟	
总体要求	《中华人民共和国宪法》第 33 条规定，国家尊重和保障人权，并规定了公民在法律面前一律平等的原则。		《欧洲人权公约（ECHR）》：任何声称其权利在被成员国侵犯的个人都可以向人权法院法院提出申诉。 《欧洲联盟基本权利宪章》和一系列非歧视指令，要求成员国将其转化为国内法。	
主要法律和指令	劳动法	禁止在就业中基于种族、性别、宗教信仰等的歧视行为。	种族平等指令	禁止在多个领域基于种族或民族出身的歧视。
	就业促进法	强调公平就业，禁止用人单位在招聘中基于性别、民族等因素进行歧视。	就业平等指令	禁止在就业领域基于宗教、信仰、残疾、年龄或性取向的歧视。
	妇女权益保障法	保障妇女在政治、经济、文化等方面的平等权利，禁止任何形式的歧视。	性别商品和服务指令	禁止在商品和服务供应领域基于性别的歧视。
	个人信息保护法	规定了个人信息处理的透明度要求，保护个人信息免受滥用，间接支持反歧视目标。	通用数据保护条例	规定了数据处理的透明度要求，防止数据处理过程中出现歧视性结果。

图 3.1 中国与欧盟反歧视立法比较

正如上图所示，中国和欧盟都在其宪法或基本权利文件中明确规定了平等和非歧视的原则，体现了两者对人权和公平正义的高度重视。双方都有具体的法律和指令来执行这些基本原则，涵盖了就业、性别平等等多个领域，并且都意识到保护个人信息对反歧视法律体系的重要性。通过制定严格的数据保护与透明度要求，防止数据处理过程中出现歧视性结果，从而增强公众对法律体系的信

任^[1]。尽管中国和欧盟在法律体系、执行机制等方面存在差异，但这些差异与本文的核心论点无太大关系。

总体而言，中国和欧盟在反歧视法规上的相似之处，为构建人工智能相关的反歧视法律体系提供了宝贵的指导意义。这些共同点不仅奠定了明确的法律基础，确保了 AI 系统设计和应用中的公平性，还在就业领域中严禁基于种族、性别、宗教等因素的歧视行为，保障了 AI 在招聘和就业决策中的公正性。此外，通过严格的数据保护与透明度要求，防止数据处理过程中出现歧视性结果，增强了公众对 AI 系统的信任。这些共同点为构建一个公正、透明且具包容性的 AI 反歧视法律框架提供了坚实的基础。综上所述，中国和欧盟在反歧视法律方面的共同努力和经验，不仅为各自国内的公平正义提供了保障，也为全球范围内人工智能反歧视法律体系的构建提供了有力的借鉴和支持^[2]。

^[1] 李开复：《人工智能：未来发展的机遇与挑战》，清华大学出版社，2019 年印行。

^[2] Craig, P., & de Búrca, G. (2015). EU Law: Text, Cases, and Materials. Oxford University Press.

四、 遏制人工智能相关的反歧视法律体系

理论上，混合的反歧视法律体系可以最大限度地发挥不同理论体系的优势，并尽可能减少其缺陷。这引发了一个问题：在面对人工智能带来的特殊挑战时，是否可以得出相同的结论？回顾文章的第二部分，我们详细解释了这些特殊挑战。简言之，我们将人工智能驱动的歧视问题分为两类：一类是伤害特定群体的差异化；另一类是不伤害特定群体但仍不公平的差异化。

（一） 完全封闭、开放的法律体系难以适用算法歧视

与完全封闭的法律体系相比，完全开放的法律体系更能适应人工智能驱动的差异化。法律的制定和更新速度远远低于科技发展的速度，因此完全封闭的法律体系在面对 AI 技术的快速发展时往往显得力不从心。这样一来，封闭的法律框架可能无法及时回应技术变革所带来的新挑战，无法有效规制 AI 决策中的差异化问题^[1]。

完全开放的法律体系也并非完美无缺。尽管开放的法律体系在面对技术进步时表现得更加灵活，但它同时也给企业带来了法律不确定性。企业在使用特征数据驱动人工智能算法时，常常会面临法律上的困惑，不知道哪些做法是合法的。这种情况下，公司可能需要在每一个人工智能驱动的差异化应用中，展示差异待遇的合理理由。然而，许多算法，尤其是深度学习中的深度神经网络，因其复杂的内部结构和大规模参数，被视为“黑箱”，即不可解释的。深度神经网络在处理大规模数据和复杂任务时表现出色，但其决策过程涉及许多层次和参数，难以直观解释。这种算法的不可解释性，给企业带来了巨大的负担，因为他们需要在每一个人工智能驱动的差异化案例中，展示差异待遇的合理理由。这不仅增加了合规成本，还可能遏制科技创新的步伐。

^[1] European Union Agency for Fundamental Rights. (2018). The EU Charter of Fundamental Rights: A Commentary. Hart Publishing.

（二） 应对算法歧视的最佳反歧视法律体系

根据上述分析，我们可以用排除法得出，在反歧视立法中采用混合法律体系是应对算法歧视的最佳方法。以下将对此进行正面论述。

首先，这种混合模式能够起到重要的基准作用，并突出非歧视法律的象征性价值。公司、公共机构和法院可以将算法生成的具体歧视理由或代理与列出的歧视理由进行比较，以查看它们是否在某种程度上相似，例如是否反映了类似的偏见、成见或过于广泛的刻板印象。如果在具体案例中存在这种相似性，算法制定者必须提出有力和令人信服的理由，以证明使用算法的公平性^[1]。

其次，混合法律体系通过判例法的发展，能够动态地扩展保护范围。法院在处理具体案件时，可以根据实际情况和社会发展的需要，逐步认定新的受保护理由。这种逐案判决的方式，使得法律体系能够及时应对新的歧视风险，从而保障公民的平等权利。例如，如果某一特定算法对低收入群体产生了不公平的影响，法院可以认定这种基于经济地位的差异化对待构成歧视，从而要求算法开发者提供有力的正当理由来证明其算法的合理性和公平性。随着科技的迅猛发展，新的算法和数据分析技术层出不穷。传统法律体系的滞后性使其难以迅速应对这些变化，而混合法律体系通过其动态发展的判例法体系，可以及时回应科技进步带来的法律挑战。

再次，采用混合法律体系也有助于增进公众对算法公平性的信任。通过确保法律不仅能够应对当前的歧视问题，也能预防未来潜在的歧视风险，可以增强社会对技术进步的接受度。公众需要确信，即使是最先进的技术，其应用也将受到公平和正义的监督。此外，混合法律体系中的透明度要求和案例审查机制可以帮助揭示算法决策的工作原理及其潜在的不公平性，从而促使算法开发者在设计和实施阶段考虑到这些法律要求^[2]。

此外，混合法律体系中的动态判例法发展机制可以促使企业不断改进其算法和技术，以适应新的法律和社会要求。例如，企业在开发新产品或服务时，必须考虑到法律判例中揭示的潜在歧视问题，并采取相应的措施来消除这些问题。这种法律驱动的技术改进不仅提高了产品和服务的质量，也增强了企业的社会

^[1] 宋乐、林德康：《人工智能：从图灵测试到深度学习》，科学出版社，2020年印行。

^[2] Craig, P., & de Búrca, G. (2015). EU Law: Text, Cases, and Materials. Oxford University Press.

责任感和品牌声誉^[1]。

（三） 混合法律体系利于与法规编码技术相结合

1. 编码法规技术

编码法规技术是一种将法规中的法律条文转化为特定应用领域的机器可读政策语言的过程。其主要目的是确保算法能够正确理解并执行特定的法规和政策，从而在数据处理过程中保持合规性。

该技术的概念起源于计算机科学与法律交叉领域的早期探索。在 20 世纪末和 21 世纪初，随着信息技术的迅猛发展，数据保护和隐私问题逐渐成为社会关注的焦点。为了应对这些挑战，学术界和工业界开始尝试将法律条文转化为可执行的机器语言，以便在数据处理和系统设计中更好地贯彻法律要求。最初的尝试主要集中在金融和医疗等高度监管的行业，但随着技术的进步，应用领域逐渐扩展到更多的行业和场景。

编码法规的核心在于将抽象的法律概念精确地映射到具体的数据处理操作和字段中。这需要多方专家的协作，包括数据保护官员、法律专家和领域专家，以确保法规翻译的准确性和可操作性。具体来说，编码法规技术包括以下几个关键步骤：

1. **法规解析**：对法律条文进行详细解析，提取其中的关键条款和要求。
2. **政策语言转换**：将解析后的法律条款转化为机器可读的政策语言。这种语言需要既能表达复杂的法律逻辑，又能被算法和系统所理解和执行。
3. **系统集成**：将转换后的政策语言集成到现有的数据处理系统中，以确保在实际操作中能够实时执行和验证法规合规性。

编码法规技术本器已经在多个部门法领域得到了广泛应用，并形成了一些主流技术和工具。例如，政策引擎如 Open Policy Agent (OPA) 通过定义灵活的政策语言，允许企业在各种应用场景中实施和执行复杂的法规要求；法律本体论如 LegalRuleML 通过创建法律概念和条款的本体模型，使得法规翻译更具系统性和一致性；自动化合规检测工具如 Privacy Policy Language (PPL) 通过自动化

^[1] Pasquale, F. (2015). *The Black Box Society: The Secret Algorithms That Control Money and Information*. Harvard University Press.

分析和检测，帮助企业在数据处理过程中实时监控和验证合规性。这些技术和工具的广泛应用，大大提升了法规执行的效率和准确性^[1]。这些技术和工具的广泛应用，大大提升了法规执行的效率和准确性。。

2. 混合的反歧视法律体系与法规编码技术结合

将混合法律体系与法规编码技术结合，可以高效执行反歧视法律并动态适应其变化，使其可以在算法执行的开始之前，通过的形式化的法规对算法进行解析，从而检测相关算法是否是歧视性的。

(1) 技术实现流程：从法律条文到机器可读政策

1. **制定并翻译基础政策：**首先，法律专家、数据保护官员和领域专家齐心协力，对相关法律条文进行深入解析并将抽象的法律概念转化为具体的、可操作的政策。这一过程中，专业知识的碰撞与融合，确保了政策的准确性和完整性。制定出的基础政策不仅要涵盖法律条文的核心要求，还需结合具体应用场景，确保其在实际操作中的可行性。
2. **编码政策语言：**通过使用如 LEGALEASE 这样的政策语言，将制定好的基础政策进行编码。LEGALEASE 是一种专门设计用于表达法律和政策条款的机器可读语言，能够精确地捕捉法律条文中的细节和逻辑关系。编码过程中，政策语言被转换成能够被计算机系统理解和执行的格式。这一过程不仅需要深厚的法律知识，还需要对计算机语言和编程的深入理解^[2]。
3. **集成到数据处理操作中：**一旦政策编码完成，这些机器可读的政策便被集成到具体的数据处理系统中。在这一阶段，确保政策能够在实际操作中得到正确执行尤为重要。系统开发者需要对数据处理流程进行详细的设计和调调整，以便将法律要求无缝嵌入到数据处理的每一个环节中。
4. **自动化合规性检查：**使用静态分析工具，如 PRIVANALYZER，实现对数据处理过程的自动化合规性检查。PRIVANALYZER 能够对系统进行深入的

^[1] Hogan, J., & Zardini, A. (2017). Open Policy Agent: Policy-based control for cloud-native environments. Proceedings of the 2017 IEEE 10th International Conference on Cloud Computing.

^[2] Surden, H. (2012). Computable Contracts. UC Davis Law Review, 46(2), 629-700.

静态分析，自动检测潜在的合规风险，并提供详细的合规性报告^[1]。

(2) 可行性分析

混合法律体系结合了封闭法律体系和开放法律体系的优点，既提供了明确的法律确定性，又保留了一定的灵活性。这种体系在与法规编码技术结合时，具有显著的优势。首先，混合法律体系中的封闭部分提供了明确的、具体的法律条文和规则，使得编码法规技术能够将这些条文准确地翻译成机器可读的政策语言。这种明确性减少了算法执行中的模糊性和不确定性，从而提高了法律执行的准确性和一致性。其次，混合法律体系的开放部分允许法律在面对新兴技术和新型歧视形式时具有灵活性。这种灵活性使得编码法规技术能够更好地适应和处理复杂的实际应用场景，通过定期更新和调整编码政策，确保算法能够及时反映最新的法律要求和社会变化。

此外，混合法律体系需要数据保护官员（DPO）、法律专家和领域专家的协作，才能将抽象的法律概念准确地映射到具体的数据处理操作和字段中。这种协作促进了不同领域专家之间的交流和合作，使得法规翻译过程更加全面和准确，确保编码法规能够全面覆盖法律要求。通过将混合法律体系中的规则和原则编码为机器可读的政策语言，数据处理过程中的合规性检查和执行可以实现自动化。这种自动化不仅提高了合规检查的效率，还减少了人工操作的错误，确保数据处理过程始终符合法律规定。

最后，混合法律体系的灵活性使得其能够支持复杂的数据处理和分析任务。在数据分析和机器学习等领域，数据处理过程往往涉及多个步骤和复杂的操作。通过将法律要求编码为机器可读的政策，能够在每一步操作中自动检查和确保合规性，从而支持复杂数据处理任务的顺利进行。

^[1] Garg, V., & Garg, D. (2015). PRIVANALYZER: A Tool for Checking Anonymity and Privacy in Data Sets. Proceedings of the 2015 IEEE International Conference on Data Science and Data Intensive Systems.

五、 总结与展望

随着人工智能技术的迅猛发展，算法歧视问题日益显现并日趋严峻。未来，立法者和执法者需要密切关注并深入研究这一领域，不断制定和完善相关法律法规，以应对瞬息万变的技术和社会环境。混合法律体系与法规编码技术的结合，展现了应对算法歧视的有效路径，但其具体实施仍需进一步探索与实践。

未来的研究应聚焦以下几个方面：首先，深入研究和开发更加先进的法规编码技术，提高其自动化合规性检查的精准度和效率；其次，推动跨学科合作，结合法律、计算机科学和社会学等多个领域的专业知识，共同应对算法歧视问题；最后，加强国际合作，分享并借鉴各国在反歧视法律体系和技术实施方面的宝贵经验和成功案例，共同推动全球范围内人工智能技术的公平和正义应用。

通过持续的法律和技术创新，以及广泛的跨领域和国际合作，我们有望构建一个更加公正、透明且包容的人工智能反歧视法律框架，为社会的可持续发展和人类的福祉提供坚实保障。这一框架不仅将推动技术进步，还将确保每一个人都能在公平的环境中受益于科技带来的便利和进步。

参考文献

- [1] Sweeney L. Discrimination in online ad delivery[J]. Communications of the ACM, 2013, 56(5): 44-54
- [2] Leese M. The new profiling: algorithms, black boxes, and the failure of anti-discriminatory safeguards in the european union[J]. Security Dialogue, 2014, 45(5): 494-511
- [3] Russell S J, Norvig P. Artificial intelligence: a modern approach[M]. Pearson, 2016
- [4] Turing A M. Computing machinery and intelligence[M]. Springer, 2009
- [5] Makkonen T. European handbook on equality data: why and how to build a national knowledge base on equality and discrimination on the grounds of racial and ethnic origin, religion and belief, disability, age and sexual orientation[M]. Office for official publications of the European communities, 2007
- [6] Eubanks V. Automating inequality: how high-tech tools profile, police, and punish the poor [M]. St. Martin's Press, 2018
- [7] Barnard C. Eu employment law[M]. OUP Oxford, 2012
- [8] Muir E, Waddington L. Directive 2000/78/ec establishing a general framework for equal treatment in employment and occupation[C]. International and European Labour Law. 2018: 520-547
- [9] Toggenburg G. Fundamental rights and the european union: how does and how should the eu agency for fundamental rights relate to the eu charter of fundamental rights?[J]. EUI Department of Law Research Paper, 2013(13)
- [10] Toggenburg G N. The european union fundamental rights agency[J]. International Human Rights Institutions, Tribunals, and Courts, 2018: 443
- [11] Craig P, De Búrca G. Eu law: text, cases, and materials uk version[M]. Oxford University Press, USA, 2020
- [12] Pasquale F. The black box society: the secret algorithms that control money and information [M]. Harvard University Press, 2015
- [13] Wang L, Khan U, Near J, *et al.* {Privguard}: privacy regulation compliance made easier[C]. 31st USENIX Security Symposium (USENIX Security 22). 2022: 3753-3770
- [14] 万国. 从图灵测试到深度学习: 人工智能 60 年[J]. 科技导报, 2016, 34(7): 8
- [15] 曹博. 算法歧视的类型界分与规制范式重构[J]. 现代法学, 2021, 043(004): 115-126
- [16] 许衍会. 算法歧视的成因, 认定与治理[J]. 2022
- [17] 李成. 人工智能歧视的法律治理[J]. 中国法学 (文摘), 2021(2)
- [18] 马皓, 宋业臻. 人工智能犯罪风险评估” 算法歧视” 现象及其规制路径[J]. 江淮论坛,

2022(2): 9

- [19] 周伟. 反歧视法研究: 立法、理论与案例: study on anti-discriminative law legislation,theory and cases[M]. 反歧视法研究: 立法、理论与案例: Study on anti-discriminative law legislation,theory, 2008
- [20] 王永庆. 人工智能原理与方法[M]. 人工智能原理与方法, 1998
- [21] 杨成铭. 《欧盟宪法条约》对欧盟人权保护的影响[J]. 法学杂志, 2006, 27(1): 4

致 谢

在本学期科技法学课程即将结束之际，怀着无比感激的心情，向刘芳老师致以最诚挚的感谢。回顾这一学期的点滴，刘老师的辛勤教诲如同阳光般温暖，照亮了我前行的道路。刘老师的教导不仅使我在科技法学的知识海洋中畅游，更在思维方式和价值观念上为我指引了方向。

每一次精彩的课堂，刘老师都以深入浅出的讲解和敏锐独到的见解，引领我透视科技与法律交汇处的复杂问题，使我深刻体会到科技进步对法律体系的挑战与机遇。这不仅使我在学术上不断进步，更教会我如何在科技飞速发展的时代，保持对法律的敬畏与对正义的追求。

在此，向刘老师致以最崇高的敬意和最真挚的感谢。感谢刘老师这一学期以来的辛勤教导，感谢老师对我的无私付出。刘老师是我心中永远的明灯，指引着我在科技法学的道路上坚定前行。最后，衷心祝愿刘老师身体健康，工作顺利，桃李满天下！

您的学生，付政烨。