

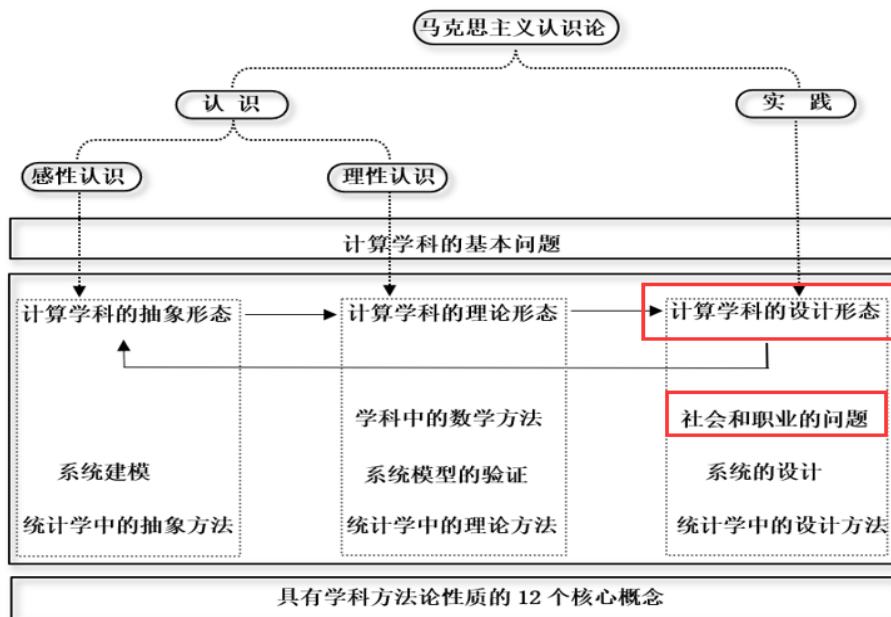
# 人工智能伦理样例：推荐系统中的大数据杀熟

古天龙 暨南大学

能力的评估：本案例能够置于 Bloom 分类法知识维度的“元认识知识”位置，学生学习后能够达到 Bloom 分类法认知过程维度的“评估”层次。

## 一、本案例课程思政的关注点

1. 本案例内容在计算学科课程思维总体框架中的位置



2. 本案例属于学科设计形态中的社会与职业的问题。算法歧视是由数据和算法分析导致的对特定群体的、系统的、可重复的、不公正的对待。平等与自由共同构成现代法治的基础，公正是现代文明国家的价值追求，算法歧视令特定群体遭遇系统性的不公正对待，在起点上违背平等的理念，在终点上导致不公正的后果，从根本上违背平等和公正的社会主义核心价值观。

3. 《中华人民共和国电子商务法》第 18 条第 1 款规定“电子商务经营者根据消费者的兴趣爱好、消费习惯等特征向其提供商品或者服务的搜索结果的，应当同时向该消费者提供不针对其个人特征的选项，尊重和平等保护消费者合法权益”，算法歧视违反了上述宪法和法律中相关条款的要求和规定。

4. 在本案例中，要求教师将 CC2020 报告中 11 个品行元素中的“责任、专业性”与该案例绑定在一起进行可操作性解释。责任，在设计人工智能的推荐系统时，要求专业人员应该有基本的责任道德意识，能够对算法歧视造成的直接或间接后果有足够的认识，不做这种有违消费者权益的事。专业性，要求学生能从专业的角度回应“大数据杀熟”对社会造成的冲击和危害，支持各类组织和个人监督平台企业的歧视性定价行为，促进平台规范、诚信经营，推动互联网平台的良性发展。

## 二、本案例的具体内容

“大数据杀熟”是推荐系统中存在的一种典型的偏见歧视方式，是指企业分析和预测消费者大数据，对不同对象收取不同价格的行为，尤其是，对于同样商品或服务，老客户得到的价格，反而高于新客户，亦即，老客户被“杀熟”。

目前“大数据杀熟”已呈泛化趋势，不局限于对熟客的“宰杀”。经营者通过大数据分析 with 用户画像，可以针对任何其所认为的“对价格不敏感”的用户实施差别定价。互联网企业掌握了大量个人信息、行为习惯等数据后，判断其偏好、用户黏合度、价格敏感度等，利用大数据技术实现“千人千面”，使不同用户看到不同的价格或搜索到不同的结果，从而被“杀熟”，这也被称为“杀熟二代”。杀熟二代中的“熟”，已经不是“熟客”，而是被平台充分掌握个人信息的“熟人”。“大数据杀熟”一般是指交易价格的差别待遇，因此“大数据杀熟”一般被定性为“差别定价”或“价格歧视”。

价格歧视（Price Discrimination）实质上是一种价格差异，通常指商品或服务的提供者在向不同的接受者提供相同等级、相同质量的商品或服务时，在接受者之间实行不同的销售价格或收费标准。经营者没有正当理由，就同一种商品或者服务，对若干买主实行不同的售价，则构成价格歧视行为。

根据差别定价的程度，英国经济学家庇古（Arthur Cecil Pigou）把价格歧视区分为三个等级：

（1）一级价格歧视（First-degree Price Discrimination）又称为完全价格歧视，即，销售者为每一位顾客及其所购买的每一单位商品制定不同的价格，因此获取所有的消费者剩余。一级价格歧视比较少见。

（2）二级价格歧视（Second-degree Price Discrimination），即，销售者对购买者偏好的多样性有所了解，但是不能观察到每一位特定顾客的特性。即便如此，仍然有可能通过提供一系列包括价格和各种条款的销售合同，在不同的购买者之间进行价格歧视。

（3）三级价格歧视（Third-degree Price Discrimination）又称为作指示选择（Selection by Indicators），即，有时企业拥有关于顾客的信息，特别是关于顾客愿意支付的价格的信息。如果购买者的特性是可以观察到的，那么销售者就可以制定不同的价格，此时，价格就可以看作购买者特性的函数。三级价格歧视是最普遍的价格歧视形式。三级价格歧视中销售者将购买者分组，对每一组制定不同的价格，这种行为也称为市场分割（Market Segmentation）。

### 1. 滴滴“苹果税”

2020年，复旦大学管理学院副教授孙某带领研究团队在五座城市打了800多趟车，花费50000多元，搜集到滴滴、曹操、首汽、美团和高德等多个渠道的数据。调查招募了20多名大学生作为调研员，对五城市（上海、北京、深圳、成都和重庆）、不同距离（近途：3公里以内，中途：3-10公里，远途：10公里以上）以及多个时间段（工作日早高峰、工作日晚高峰、日间非高峰、晚间非高峰）进行了抽样调查。获得总样本836个，有效样本821个，通过数据分析，最终形成“2020打车报告”。

该研究团队表示，其在研究过程中验证了“苹果税”的存在。团队用“一键呼叫‘经济型+舒适型’两档后被舒适型车辆接走的订单比”来判断“被舒适”的程度，数据表明，与非苹果手机用户相比，苹果手机用户更容易被舒适型车辆（如，专车、优享等）司机接单，这一比例是非苹果手机用户的3倍。

除了手机品牌，平台也可能同时关注乘客手机价格。如果乘客使用的是苹果手机，那么就更容易被推荐舒适型车辆；如果乘客不是用苹果手机，那么就要看其手机价位，手机价位越高则越有可能被舒适型车辆接走。

此外，“苹果税”还体现在打车优惠上。数据表明，苹果手机用户平均只能获得2.07元的优惠，显著低于非苹果用户的4.12元。除绝对金额外，优惠折扣比同样支持上述结论。

早在2018年，针对滴滴大数据杀熟的质疑，滴滴总裁柳青在微博澄清表示，因为路况是动态变化的，前一秒和后一秒都可能发生很大变化。为了给乘客尽可能准确的预估参考价，滴滴是实时根据动态路况预测预估价的，即便起终点一样，预估价也会存在差异。并且，两个用户的预估路径规划可能不同，对经常打某一路线的用户会用他平时常用的路径缺省值，没走过这个路线的用户系统会推荐最优方案。

## 2. 携程“钻石会员”服务

胡女士是携程的“钻石会员”，2020年7月，胡女士通过携程APP订购了舟山希尔顿酒店的一间豪华湖景大床房，支付价款2889元。离开酒店时，胡女士发现酒店的实际挂牌价仅为1377.63元。

胡女士认为，自己不仅没有享受到星级客户的优惠，反而多支付了一倍的房价。退房后，她与携程沟通，携程以其系平台方，并非涉案订单的合同相对方等为由，仅退还了部分差价。胡女士不服携程的处理结果，以携程公司采集其个人非必要信息，进行“大数据杀熟”等为由起诉到法院，要求“退一赔三和携程APP为她增加不同意《服务协议》和《隐私政策》时仍可继续使用的选项，以避免携程公司采集她个人信息，掌握她的数据。”

绍兴柯桥法院审理此案后认为：携程APP作为中介平台，对酒店实时房价有如实报告的义务，但其却未如实报告。携程向原告胡女士承诺钻石贵宾享有优惠价，事实上却没有价格监管措施，向原告展现了一个溢价100%的失实价格，未践行承诺。携程在处理原告投诉时告知原告无法退还全部差价的理由，经调查也与事实不符，存在欺骗。

法院认定，被告携程公司存在虚假宣传、价格欺诈和欺骗行为，支持原告“退一赔三”的诉讼请求。而且，携程APP收集的用户信息超越了形成订单必需的要害信息，属于非必要信息的采集和使用，其中用户信息分享给携程可随意界定的关联公司、业务合作伙伴进行进一步商业利用更是既无必要性，又无限增加用户个人信息使用的风险。

据此，法院当庭作出宣判，判决被告携程公司赔偿原告胡女士投诉后携程未完全赔付的差价243.37元及订房差价1511.37元的三倍，总计支付赔偿金4777.48元，且应在其运营的携程旅行APP中为原告增加不同意其现有《服务协议》和《隐私政策》仍可继续使用的选项，或者为原告修订携程旅行APP的《服务协议》和《隐私政策》，去除对用户非必要信息采集和使用的相关内容，修订版本需经法院审定同意。

2021年7月19日，携程就此次判决回应21世纪经济报道记者，胡女士遭遇的价格差系供应商囤房再利用库存紧张的时间点进行超高价销售，并非“价格杀熟”。此外，携程在对外情况说明中提到，公司已经认识到仍需强化对供应商“囤房加价”行为的监管拦截，并上线新版风险控制系统。

## 3. 美团“割韭菜”

2020年12月，一篇名为《我被美团会员割了韭菜》的自媒体文章成为全网爆款，再次把大数据杀熟的问题推上风口浪尖，美团成为众矢之的。名叫“漂移神父”的网友爆料称，他在美团App购买驴肉火烧外卖时发现，在同一家店铺、同一个配送位置、同一个下单时间点的条件下，开通美团会员的账号所需的配送费，比未开通会员的账号所需的配送费高出整整4元。该网友多次亲身测试后指出，其所在位置附近几乎所有外卖商户，会员账号总是超出非会员账号1-5元不等的配送费。

在反馈客服后，对方表示愿意补偿十元红包。电话客服主管则表示已超出他的处理范围，会向更高级管理人员反映。文章中还提到，不仅是一家店有这种情况，一部开通美团外卖会员的手机，附近几乎所有外卖商户的配送费，基本都要超出非会员配送费1-5元不等。换句话说，会员并未享受到平台所承诺的优惠待遇，只要多点几次外卖，额外支付的配送费甚至会超过会员所享的优惠。

针对用户反馈的“会员和非会员配送费差异”问题，美团外卖公开回应称，文中提到的配送费差异与会员身份无关，是软件的定位缓存偏差导致。用户实际下单的配送费，会按照真实配送地址准确计算，不受影响。

“美团的解释我并不认同。”文章作者在接受媒体采访时表示，第一次地址是有问题，但第二次地址是一样的，因此他不认同美团的回应。此后，文章作者再发微博表示，本次事件已然会改变大家对互联网巨头的认知，也可能会促成巨头更合规的运营。

## 三、教学体会

在算法背景下讨论歧视问题，不仅需要研究到底存在何种歧视形式，更为重要的是阐明为什么要禁止它们。从伦理的角度而言，反对算法歧视的原因在于它涉及到公平与算法身份两大伦理问题。

## 1. 算法公平

首先，算法歧视涉及到算法语境下对“公平”（Fairness）的诠释问题。学者对公平有诸多不同的理解，但在算法语境下讨论公平涉及到两个问题：其一，公平是否可以被量化、形式化？其二，如果公平可被量化、形式化，那么选择何种公平理论是恰当的？

对于第一个问题，从当前各领域的研究成果来看，答案是肯定的。防止歧视的数据挖掘和算法公平俨然成为新兴的研究范式，试图借此发现并消除算法引发的不公平现象。康奈尔大学巴罗卡斯（Solon Barocas）、普林斯顿大学克罗尔（Joshua Kroll）等人建立了专注于机器学习公平、透明与可责的研究中心（Fairness, Accountability, and Transparency in Machine Learning, FAT-ML），鼓励计算机领域的学者在预防算法歧视的前提下设计算法。

针对第二个问题，公平在算法中通常被量化、形式化为两种形式：群组公平（Group Fairness）与个体公平（Individual Fairness）。群组公平又被称为统计奇偶性（Statistical Parity），指的是接受正分类或负分类的人的比例与整个人口统计是相同的，旨在平等地对待所有群体，它要求通过算法而进行的决策结果在受保护群体与非受保护群体之间的比例相等。例如，若某行业从业人口的男女比例为 1:2，那么受雇佣的男女比例也应接近 1:2。然而，德沃克（Cynthia Dwork）等人则认为群组公平并不足以保证算法决策结果的公平性，因为就个人角度而言，它可能造成不公平。例如，在申请贷款中，群组公平并不能保证两个群体中信誉度相似的两个人获得相同的贷款概率。因此，他们提倡另一种形式的公平——个体公平。

个体公平是平等地对待相似的个体。德沃克通过测量在特定任务中个体之间的相似度而保证分类的公平性，以防止个体由于从属于某一群体成员而被歧视。约瑟夫（Matthew Joseph）则将个体公平进一步发展为个体机会均等：通过将罗尔斯（John Bordley Rawls）的《正义论》中的“机会平等原则”以数学的方式形式化，建立机器学习中的上下文赌博机（Contextual Bandits），以保证算法的每一步都公平地对待相似的个体。

事实上，在“群体公平”与“个体公平”选择的背后隐藏着设计者的不同世界观。“群体公平所隐含的是‘我们都是平等的’，即，在算法的特征选择过程中，个体所组成的群组之间并没有天生的差异，他们看起来都是一样的。或者说，即使存在差异，这些差异都是由个人控制之外的因素造成的，在算法决策过程中不应考虑这些因素。个体公平隐含的则是‘所见即所得’。也就是说，针对算法特征选取的观察方法能够较为真实地、准确地表征个体并作为算法的输入。”

算法语境下公平的形式化仅是对公平阐释的内容之一，而非全部。对此，卡内基梅隆大学计算机伦理学家斯克潘（Michael Skirpan）认为，算法语境下对公平的诠释需要重视两个问题：

“一是诠释算法公平时，公平更多的是一种公平建构（Fairness Construct），即公平不是静态的，而是动态的，对其诠释依赖于算法所要解决的特定问题空间以及在训练和应用中对该诠释进行编码和测量的方法。”

“二是需超越上述对技术方法之公平的关注，重视算法系统的公平性与结果的公平性。算法系统的公平性即采用算法系统解决某个特定问题是否恰当。例如，算法驱动的路径导航的自主实验系统不仅在缺乏用户知情同意的情况下开展实验，而且在实验过程中整个算法系统可能会更倾向于选择那些信息较少的用户作为实验对象，这就提出了如何在实验对象与非实验对象之间公平分配利益与风险的问题——结果的公平性，即，算法的结果是否公平。它可被理解为对算法歧视的事后控制，可以使用特殊的事后分析方法发现黑盒模型可能存在的不公平。”

## 2. 算法身份

第一，算法歧视关涉个体被赋予何种算法身份。在算法决策中，个体被赋予一种新的身份，即，算法身份。但它并非静态的、不可更改的，而是具有交互性、适应性和可跟踪性等特征：





## 五、 习题

1.什么是大数据杀熟？你对其治理有何建议？

### 参考文献

- 1.古天龙.人工智能伦理导论[M].高等教育出版社, 2022
- 2.古天龙,李龙,常亮,罗义琴.公平机器学习:概念、分析与设计[J].计算机学报,2022,45(05):1018-1051
- 3.Xuan Feng, Tianlong Gu, Xuguang Bao, Long Li. Behavior-based Ethical Understanding in Chinese Social News. IEEE Transactions on Affective Computing, doi:10.1109/ TAFFC. 2022. 3160745
- 4.常亮,张伟涛,古天龙,孙文平,宾辰忠.知识图谱的推荐系统综述[J].智能系统学报,2019,14(02):207-216