arXiv:2601.18654v1 [cs.CY] 26 Jan 2026

何时自我披露最优？人工智能生成内容的激励与治理

Juan Wu[1], Zhe (James) Zhang [2], Amit Mehra[2]

[1] 中国科学技术大学 [2]达拉斯德克萨斯大学

摘要

生成式人工智能（生成式人工智能）通过降低生产成本并实现不同质量内容的规模化产出，正在重塑数字平台上的内容创作。对此，平台开始采用披露政策，要求创作者标注人工智能生成内容，并通常辅以不完善的检测和对不合规行为的惩罚。本文建立了一个形式化模型，研究此类披露制度的经济影响。我们比较了两种情形：在不披露基准下，只有平台检测人工智能使用；在强制自我披露制度下，创作者在执法不完善的情况下策略性地选择披露还是隐瞒人工智能使用。模型纳入了异质创作者、观看者对人工智能标注内容的折价、检测到不披露后的信任惩罚，以及内生执法。分析表明，只有当人工智能生成内容的价值和其节省成本优势均处于中等水平时，披露才是最优的。随着人工智能能力提升，平台的最优执法策略会从严格威慑转向部分筛查，最终走向放松管制。尽管披露能够可靠地提高透明度，却会降低创作者总剩余；当人工智能技术较为先进时，还可能抑制高质量AI内容。总体而言，研究结果表明，披露是一种战略性治理工具，其有效性取决于技术成熟度和信任摩擦。

1 引言

生成式人工智能（Gen-AI）正在迅速改变数字内容平台上的创意生产。近年来，大规模生成模型（如 GPT-5.2 和 Gemini Ultra）取得的进展，显著拓展了内容创作的范围、速度和可扩展性。这些技术使创作者能够以前所未有的规模制作文本、图像、音频和视频，同时大幅降低

生产成本。对于平台而言，生成式人工智能的普及有望带来更大规模、更多样化的内容供给，降低创作者的进入壁垒，并通过提升参与度和广告收入，扩大变现机会。

与此同时，生成式人工智能的迅速普及给内容平台带来了根本性的治理挑战。人工智能工具虽然能够提升生产力和内容质量，但也引发了人们对真实性、可信度和用户信任的担忧。既有研究表明，受众往往认为人工生成内容更真实、更能引发情感共鸣，因此参与度和情感依附也更高（Wanet al. 2017、Kim andDennis 2019）。然而，随着人工智能生成内容越来越像人类创作，区分人类创作与机器创作在技术上日益困难，在策略层面也愈发重要。观看者可能重视内容创作过程的透明度，但这种透明度也可能引发怀疑并降低参与度，尤其是在对可信度要求较高的领域。

过去，平台一直依赖算法检测来识别人工智能生成或经过操纵的内容，并利用语言模式、元数据和行为信号进行判断。然而，检测准确率仍不完美，且维护成本高昂。实证证据表明，最先进的检测器准确率往往低于 80%，且在不同语言、格式和对抗性释义改写情形下表现不一（Weber-Wulff et al. 2023、Sadasivan et al. 2023、Kalpesh etal.2023）。检测模型还需要持续重新训练并使用标注数据，因此无法成为大规模应用的完整解决方案，甚至可能导致巨额成本。此外，仅靠检测无法解决创作者的策略性激励问题：创作者可能会根据预期的受众反应和平台执行情况，选择披露或隐瞒 AI 使用情况。

对此，平台越来越多地转向要求创作者自行报告人工智能使用情况的披露政策。自2024年初以来，Meta已强制要求在Facebook和Instagram上使用“由人工智能构想”等标签，而YouTube则对人工智能生成或操纵的媒体实施了“修改或合成内容”披露要求。[1] 这些政策通常将自愿自我披露与作为兜底机制的算法检测相结合，并在检测到未披露的人工智能使用情况时施加惩罚，例如降低可见度、降低排名或移除内容。[2] 然而，各平台的披露实践差异很大，而且一些主要平台（如 X、Reddit 和 Bluesky）尚未采用强制披露制度。

尽管披露政策日益普遍，但其对人工智能生成内容的经济后果仍鲜为人知。披露通常被视为一种伦理干预或提升透明度的举措，但它也作为一种战略性治理工具发挥作用，重塑创作者激励、内容

[1] https://about.fb.com/news/2024/02/labeling-ai-generated-images-on-facebook-instagram-and-threads/; https://blog.youtube/news-and-events/disclosing-ai-generated-content/ [2] https://about.fb.com/news/2024/04/metas-approach-to-labeling-ai-generated-content-and-manipulated-media/和 https://transparency.meta.com/policies/community-standards/misinformation。

构成和平台价值。披露会改变人工智能采用如何转化为参与度和收入，因此影响的不仅是创作者是否使用人工智能，也包括他们选择如何以及何时披露这种使用情况。相同的人工智能生成内容，可能因人工智能使用情况是由创作者自愿披露、经平台检测后揭示，还是被策略性隐瞒，而产生截然不同的参与度结果。因此，披露不仅改变采用的边际决策，还从根本上改变创作者在透明度、隐瞒和内容定位方面的战略权衡。从这个意义上说，披露是一种内生激励机制，在不同创作者类型和内容领域之间重新分配人工智能使用，而非对技术采用施加简单的二元限制。

本文探讨了一系列相互关联的研究问题，涉及数字平台上人工智能生成内容的治理。

1. 在什么条件下，平台应强制披露人工智能使用情况，而非仅依赖算法检测？

2. 披露政策如何重塑创作者的策略性激励？这不仅涉及是否采用生成式人工智能，也涉及是否披露、隐瞒或选择性透露其使用情况。

3. 随着生成式人工智能技术的质量和成本效率不断变化，平台应如何优化执法力度，例如针对不披露行为设定何种惩罚？

4. 披露制度如何影响平台盈利能力、创作者剩余和总体内容质量？透明度提升又会在何时以生态系统价值下降为代价？

通过在统一的形式化框架内回答这些问题，我们阐明了披露政策何时能发挥有效治理工具的作用，何时反而会成为扭曲人工智能采用和价值创造的阻力。具体而言，我们构建了一个形式化模型，对比两种制度：在不披露基准下，平台独自检测并标记人工智能生成内容；在自我披露制度下，创作者必须标注人工智能使用情况，并接受不完全检测以及对不合规行为的惩罚。该模型涵盖了生成式人工智能生态系统的三个关键特征。第一，人工智能能够降低生产成本，并可能提升内容质量，但这些影响的程度因创作者而异。第二，观看者会降低标注为人工智能生成的内容的评价，并进一步惩罚被平台发现未披露的创作者，这反映了算法厌恶和信任顾虑（Castelo等（2019），Logg 等（2019），Longoni等（2019））。第三，检测并不完美，因此会产生隐瞒的策略性动机。

该分析得出了若干重要结论。只有在一个中间区域内，披露才是最优的；在该区域中，人工智能生成内容具有价值但并不占据主导地位，成本节约显著但并非压倒性。在这一地区，披露提高了透明度，并以一种能够维护信任、同时不会严重扭曲生产效率的方式，在不同创作者类型之间重新配置人工智能使用。在该区域之外，披露会变得

适得其反。当人工智能质量较低时，披露主要会抑制采用，却不会提高内容价值。当人工智能质量较高时，披露引致的可信度折扣和执行成本会扭曲高效的人工智能采用，并降低总体内容质量。

结果进一步表明，披露对不同市场参与者并非中性。尽管自我披露能够稳定提升透明度，并可通过缓解信任外部性提高平台利润，但代价是将合规成本和可信度成本转嫁给创作者。因此，即使在该政策对平台而言具有私人最优性的情形下，披露仍会降低创作者总剩余。这种分歧凸显了披露是一种战略性治理工具，而非中性的透明度机制，并引发了人们对基于披露的治理在人工智能密集型内容市场中长期可持续性的疑问。

本文从三个方面对平台治理和人工智能赋能的生产相关文献作出贡献。第一，本文将披露建模为一种内生激励机制，它会影响创作者的采用、披露和隐瞒决策，而非被视为被动的透明度要求。第二，本文结合不完全检测、观众折价以及创作者可信度的异质性，以刻画信任摩擦如何与人工智能内容市场中的技术效率相互作用。第三，通过联合分析平台政策设计和创作者行为，本文提供了一个统一的经济学视角，将披露视为一种在创作者和内容领域之间重新分配人工智能使用的机制，并将现有的信息披露和平台监管模型拓展至生成式人工智能情境。

本文其余部分安排如下。第2节回顾相关文献。第3节介绍模型。第4节分析披露制度和不披露制度下的均衡结果，并刻画平台的最优政策。第5节讨论管理和政策启示，并作出结论。

2 文献综述

2.1 平台对人工智能生成内容的响应

针对人工智能生成内容和经过操纵的内容日益增多的现象，数字平台采取了多种治理策略，涵盖从经济优化到算法执法等方面。从经济和行为视角来看，Papanastasiou和Yiangos（2020年）构建了一个社会学习框架，用于考察平台如何根据环境波动性和激励结构动态调整检测机制，从而为不确定性下的策略性内容审核提供了见解。在此基础上，Liu等人（2022年）研究了不同变现模式下内容审核决策的差异，发现对于广告型平台而言，提高内容审核强度并不一定

会带来最优的平台结果。在这一框架基础上，Jiménez-Durán （2023）在Twitter上开展了两项现场实验，发现仇恨言论审核会对在线话语和线下仇恨犯罪产生抑制作用，凸显了社会效益与平台参与度之间的权衡。这种经济张力在（Madio和Quinn 2024）的研究中得到了进一步探讨，他们考察了平台如何在广告商对安全环境的偏好与用户对侵扰性广告的容忍度之间寻求平衡。他们的分析表明，尽管更严格的内容审核可能提高广告商满意度，但也可能无意中增加用户接触广告的程度，从而降低用户体验。

与这种战略视角相辅相成的是，人们提出了多种技术解决方案，以检测和缓解遭到篡改或合成的内容。Cavusoglu 等人(2009)研究了防火墙与入侵检测系统（入侵检测系统）之间的相互作用，说明根据配置不同，这些系统既可以互为补充，也可以相互替代。聚焦于虚假信息，Pérez-Rosas 等人(2017)构建了机器学习模型，在假新闻检测中实现了 76% 的准确率，为开发可扩展的分类工具做出了贡献。最近，Kirchenbauer 等人(2023)为专有语言模型提出了一种水印框架，揭示了在人工智能输出中嵌入可追踪信号的潜力与局限。同样，Tiwari 等人(2024)利用半监督生成对抗网络架构检测合成视频。最后，Sietsema 等人(2023)探讨了面向用户的信息披露策略，例如采用营养标签式标注来说明新闻真实性。他们的研究结果表明，这些标签的有效性并非固有属性，而是受到党派倾向、政治意识形态和认知复杂性等个体层面特征的影响。此外，未标注的错误信息并不会因此获得可信度提升，这表明标注方案的影响可能有限。

这些研究揭示了平台应对人工智能生成内容的多维特性。它们凸显了内容审核强度、用户参与度、广告商价值与技术可行性之间的战略权衡，说明在人工智能驱动的内容生态中，需要建立全面的治理框架，使平台利益与更广泛的公共福祉相协调。

2.2 AI使用激励

大型语言模型（LLM）的快速发展正在从根本上改变人机协作，尤其是在创意任务和知识密集型任务中。现有文献探讨了大型语言模型不断发展的能力，以及它们对不同协作模式下用户表现的影响。

大语言模型已展现出解决各种复杂问题的卓越能力。Noy 和 Zhang (2023)证明，人工智能可能不会取代专业人士，而是作为一种工具，提升整体生产力，尤其是帮助技能较弱的群体表现得更好—

—尤其是帮助技能较弱的群体表现得更好。Bubeck 等人 (2023)强调，GPT-4 无需特殊提示，就能处理数学、编程、视觉、医学、法律、心理学及其他领域中的新颖且具有挑战性的问题。Chen 和 Chan(2024) 进一步拓展了这一观点，证明这些模型能够协助用户完成广泛的开放性任务（如问答、客户支持、摘要）和创造性任务（包括诗歌创作、歌曲创作和营销内容创作）。Eloundou 等人 (2023)的研究表明，在美国，借助大语言模型，约 15% 的劳动者任务可以在保持相同质量水平的情况下显著更快地完成。若将基于大语言模型构建的软件和工具纳入其中，这一比例可能提高至 47% 至 56%。

尽管大语言模型能力出众，但在创造力方面仍有局限。多项研究致力于提升大语言模型的创造性表现。例如，Bunescu and Uduehi(2019)提出了一种双模型架构，其中受众模型用于学习预期，并与创作者模型相连，后者用于学习如何带来惊喜，旨在提升人工智能系统的创造能力。Ouyang 等人(2022年)强调了大语言模型的指令遵循能力，并提出了一种途径：通过基于人类反馈进行微调，使语言模型能够在广泛任务中与用户意图保持一致。Bunescu and Uduehi (2019)指出，由于大语言模型具有自回归特性，它们可能受限于生成预期的输出，因为训练过程使其遵循现有的数据分布，缺少“惊喜”元素。Chen and Chan (2024)进一步强调，让大语言模型主导创作过程或许并非最佳选择。当创造力是输出质量的关键要素时，锚定效应可能尤为明显，从而限制最终输出的创造力。仅依赖大语言模型生成内容会束缚创造力，对专家用户而言尤其如此。

大语言模型为人机协作带来了新的可能性，在提升工作效率和帮助非专业用户方面具有显著优势。然而，在创造性任务中依赖大语言模型时仍需谨慎，尤其是对专业人士而言。理想的协作模式可能是将大语言模型用作辅助工具，而非完全取代人类创作过程。未来研究应着重优化人机协作方式，在发挥大语言模型优势的同时弥补其创造力方面的局限。

2.3 用户行为与平台动态

近期关于人工智能生成内容的研究揭示了内容来源、用户感知与数字平台上的行为之间复杂的相互作用。对算法决策的早期研究表明，用户最初会表现出怀疑和抵制（Dietvorst等人，2015，Castelo等（2019）、Longoni等（2019）、Liu等人

2023），而其他研究则表明，在特定条件下，用户会欣赏算法生成的结果，尤其是当这些结果超出他们的预期时（Logg 等（2019））。然而，生成式人工智能技术的应用引发了人们对偏见和真实性的新担忧。 Fang等人（2024）提供了实证证据，表明领先的大型语言模型（大语言模型）生成的内容（如 ChatGPT 和 LLaMA）存在显著的性别偏见和种族偏见，加剧了公众对公平性和代表性的担忧。

这些担忧同样涉及数字平台上观看者和创作者的行为。在观看者方面，Luo 等人(2019)进行了一项实验，结果表明，在披露之前，用户对人类销售代理和聊天机器人销售代理的反应相似；但一旦聊天机器人的身份被揭示，购买率便显著下降。这表明，感知真实性在塑造用户行为方面发挥着关键作用。Turel 和 Kalhan (2023)进一步指出，用户可能会下意识地将人工智能系统视为异类实体，从而产生偏见反应。支持这一观点的是，Brüns 和 Meißner(2024)通过一系列实验研究发现，生成式人工智能披露会引发消费者对品牌的负面态度和行为反应，这主要是因为感知真实性下降；尤其是在生成式人工智能完全取代人力进行内容创作时，这种情况更为明显。Wang 等人(2025年)基于一项在线实验发现，披露会违背消费者预期并引发负面情绪。相比之下，Schanke 等人(2024)发现，披露并不总会降低信任，尤其是在语音克隆等情境中，功能上的益处可能会盖过披露带来的影响。同样，Yin 等人(2024)表明，尽管人工智能生成的消息可以增强用户感到自己受到倾听的感受，但当用户得知消息是机器生成时，这种效果会减弱。

观众的行为也受到平台设计和内容形式的影响。Wang 等人(2022年)发现，与视频类内容相比，用户更有可能举报文本类假新闻，这表明内容模态会影响可信度评估和举报行为。平台变现机制也会改变参与度：Ye 和 Chua (2022年)表明，引入付费问答功能会同时促进内容生产和用户互动，反映出经济激励和互动模式如何塑造观众参与。

从创作者的角度来看，身份和激励会影响内容传播和策略行为。Han 等人(2020年证明，创作者属性与内容特征之间的相互作用能够显著提升内容扩散预测的准确性。Chen 等人(2019发现，货币激励会增加内容的数量和多样性，但未必能提升内容质量，这一点在金融咨询场景中尤为明显。Bhargava (2022年指出，平台提供的工具（如广告管理系统和收入分成模式）会影响创作者之间市场权力的集中程度。Ye 和 Chua (2022年还表明，问答系统的效益取决于内容应答者的身份，凸显了

创作者影响力的层级性。

人类与人工智能之间的协作方式为用户互动增添了另一层复杂性。Fügener等人(2022年) 证明，当人工智能系统将任务委派给人类时，团队绩效会有所提升；而反向委派则不会带来显著益处。这一发现凸显了人机协作中角色定位的重要性。Brüns和Meißner(2024) 发现，将生成式人工智能定位为助手而非替代者，有助于减轻用户的负面感知。Chen和Chan(2024) 表明，大语言模型作为共同创作者时，能够提升用户生成内容的质量；但作为代笔者时，则可能降低用户满意度。这些影响因用户专业水平而异；大语言模型在弥合新手与专家之间的技能差距方面尤其有益。令人欣慰的是，Yao等人(2024) 的研究表明，人工与人工智能生成的内容之间可以实现稳定均衡。然而，平台如何在不同的生成式人工智能应用场景中实现这种均衡，仍是未来研究有待解答的问题。这些研究揭示，用户对生成式人工智能的行为受到内容来源、披露、任务委派和平台设计等因素的共同影响。感知真实性、身份信号传递和激励机制，是理解人工智能介导的内容生态系统中的参与度、信任和经济成果的关键。

3 模型设定

3.1 平台、创作者与收益分成

我们考虑一个内容平台（例如 YouTube 或 Facebook），该平台拥有连续统的内容创作者。创作者发布图像、音频或视频等数字内容，平台则通过广告、订阅、捐赠及相关机制实现其货币化。一项内容产生的总收入由平台与创作者分成：平台作为佣金保留 $r \in (0,1)$的比例，而创作者获得剩余比例 $1-r$。在实践中，Meta 和 YouTube 等大型平台通常收取 30% 至 45% 的佣金率。

此外，每位创作者只创作一单位内容，其收入与观众参与度成正比。我们假设观众参与度随内容质量的提高而增加，这反映了平台变现机制由互动强度和注意力驱动这一事实。因此，我们将收入建模为内容质量的约化形式递增函数，在保留质量与变现之间核心经济联系的同时，抽象掉参与度形成的微观层面动态。表 1 总结了模型中使用的主要符号。

表1：主要符号

符号	定义
v	人工智能生成内容的质量
c	人工创作内容的生产成本
δ	人工智能生成内容的成本降低系数，$\delta \in (0, 1)$
β	平台算法的检测准确率
r	平台佣金率
f	观众对标注为人工智能生成内容的可信度折扣
k	发现未披露后的信任折扣
p	未披露人工智能使用情况的处罚
Π_i	政策 $i \in \{N, D$ 下的平台利润

3.2 内容创作技术

创作者在两种生产技术之间进行选择：人工生成内容（NAI）和人工智能生成内容（AI）。

人工生成内容需要生产成本 $c > 0$，并可达到等于一的标准化质量。人工智能生成内容能够降低生产成本，但可能改变内容质量。具体而言，使用人工智能会产生较低的成本 δc，其中 $\delta \in (0,1)$ 表示生成式人工智能替代人工劳动的程度。

我们用 v 表示人工智能生成内容的质量，其中 $v \in (0, \bar{v})$ 且 $\bar{v} > 1$。这一表述体现了这样一种观点：根据所使用的模型和应用场景，生成式人工智能的表现可能优于或逊于人类创作。更高的内容质量会带来更高的观众参与度，从而增加平台上的变现机会。当 $v > 1$ 时，人工智能生成内容产生的总参与度高于人工生成内容；当 $v < 1$ 时，人类创作在质量上占优。

3.3 平台审核与披露政策

平台使用自动检测算法审核内容，试图识别人工智能生成内容。我们考虑两种审核制度。

在不披露制度($i = N$)下，平台完全依赖其内部检测系统来标记人工智能生成内容。创作者无需披露AI使用。

在 自我披露制度 ($i = D$)下，创作者必须如实标记人工智能生成内容。如果创作者未进行披露，而平台随后检测到AI使用，平台将实施处罚$p \geq 0$，可能包括移除、降低排名、取消变现资格或账户级制裁。

检测算法以概率 $\beta \in \left(\frac{1}{2}, 1\right)$ 正确分类内容。为简化起见，我们假设检测准确率具有对称性：人工智能生成内容和人工生成内容均以概率 β 被正确识别。这一假设使我们能够将分类偏差排除在外，专注于

披露要求所引发的平台机制设计问题。我们在不同制度下保持 β 固定，以便将披露政策的激励效应与检测性能的技术差异区分开来。

为排除对AI内容的内在歧视影响，我们将不披露制度下对检测到的AI内容施加的任何直接处罚标准化为零。

3.4 观看者认知与对人工智能的厌恶

观看者从消费内容中获得效用，但可能会降低对明确标注为人工智能生成的内容的评价。这种折扣反映了对真实性、偏见和问责的顾虑，并且因观看者和内容领域而异。

我们通过可信度折扣因子 $f \in [0,1]$ 来捕捉这种异质性，该因子服从均匀分布。这一设定简洁地刻画了观众对人工智能生成内容接受度的异质性，同时确保分析上的透明度。我们的结果依赖于异质性的存在，而非分布的具体函数形式。当内容被标注为人工智能生成时，观看者对其估值为其内在质量 v 的 f。f 值越低，表示对人工智能生成内容的厌恶越强。无论披露是自愿进行还是由系统检测后披露，只要人工智能使用情况被揭示，这种感知折扣就会适用。

3.5 可信度与披露合规

披露政策也可能影响观众对创作者可信度的看法。当创作者未披露人工智能使用情况，随后又被平台标记时，观众可能会认为其有意隐瞒，从而造成声誉损害。

我们通过信任折扣因子 $k \in (0,1)$来体现这一影响。未披露后若检测到人工智能使用情况，观众会以额外的乘数因子 k 折减内容价值。若如实披露人工智能使用情况，则不会受到这种信任惩罚。

3.6 创作者收益

创作者通过发布内容来最大化期望效用，该效用取决于扣除平台佣金和制作成本后的收入。

不披露制度。在不披露制度下，创作者可以选择创作人工生成内容或人工智能生成内容。期望效用如下：

$$\begin{aligned} u_{NAI} &= 1 - r - c, \\ u_{AI} &= (1 - r)\big[\beta f v + (1 - \beta) v\big] - \delta c. \end{aligned} \tag{1}$$

当人工智能生成内容被检测到时，其价值会根据 f 打折；未被检测到时，则仅根据质量 v 进行评估。假设人工生成内容在申诉后能被完美识别，并将这一过程的成本标准化为零。

自我披露制度。 在强制披露下，创作者有三种策略：不使用人工智能（NAI）、使用人工智能但不披露（$AINAI$），或使用人工智能并如实披露（$AIAI$）。期望效用为

$$\begin{aligned} u_{NAI} &= 1 - r - c, \\ u_{AINAI} &= (1 - r)\big[\beta(f k v - p) + (1 - \beta) v\big] - \delta c, \\ u_{AIAI} &= (1 - r) f v - \delta c. \end{aligned} \tag{2}$$

相较于不披露制度，在披露制度下，一旦不合规行为被检测到，创作者不仅会遭受声誉损失（k），还会受到惩罚（p）。如实披露可以避免这些损失，但创作者必然会遭受观众对人工智能生成内容的歧视。

我们将讨论范围限定在一个非平凡的参数区域：在不存在披露摩擦时，人工智能生成内容具有经济可行性，但无论采取何种政策，它都不会始终优于人类创作，从而确保披露激励具有实际意义。具体而言，我们假设

$$\underline{v}_1 \equiv \frac{1 - r - c(1 - \delta)}{1 - r} < v.$$

如果 $\underline{v} < v_1$，创作者将永远不会采用人工智能，披露政策也无关紧要：所有均衡结果都会归结为仅有人类创作内容的情形。

我们将披露制度视为平台在内容创作之前选择的一项政策。为阐明经济机制，我们首先分别求解每种制度下的均衡，然后比较平台收益，以刻画最优政策选择。

3.7 时间线

事件的先后顺序如下：

1. **平台决策**。 平台选择是否实施自我披露政策。若实施，则设定对不合规行为的处罚 p 。

2. **创作者决策**。 创作者选择生产技术；在实行披露政策时，还要决定是否如实披露AI使用情况。

3. **观看者回应**。 观看者观察内容和标签，对内容质量及创作者可信度形成认知，并决定参与程度。

4 均衡分析

我们通过逆向归纳法刻画子博弈完美均衡。在第3阶段，观看者根据观察到的内容和标签作出回应，按照可信度折扣 f 评价人工智能标注内容，并在适用时考虑信任折扣 k。在第2阶段，创作者选择创作技术；在强制披露制度下，还要决定是否如实披露AI使用情况。在第1阶段，平台选择审核制度；在实行披露政策时，还要设定对被发现的不合规行为的处罚 p 。

4.1 基准：不披露制度

在不披露的情况下，创作者通过比较方程（ NAI 和 AI ）中的效用，在1之间做出选择。由于只有在标注时，AI内容才会被打折，创作者激励取决于接受度的抽样结果 f：真实性敏感领域（低 f）的创作者较不愿意采用人工智能，而对人工智能接受度较高的领域（高 f）的创作者则更愿意采用。

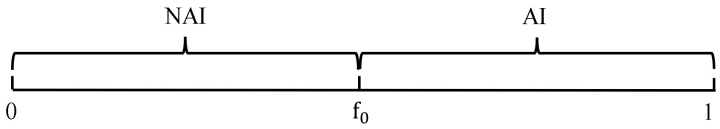

图1：不披露政策下创作者使用生成式人工智能的选择

创作者临界值与细分。当人工智能质量处于中等水平时， $v_1 < v <\hat{} v$，均衡呈现一个临界值 $f_0 \in (0,\ 1)$，使得满足 $f < f_0$ 的创作者选择 NAI，而满足 $f \geq f_0$ 的创作者选择 AI。这一

这种排序如图1所示。形式上，

$$f_0 = \frac{(1-r)\big(1-v(1-\beta)\big) - c(1-\delta)}{(1-r)v\beta}, \qquad \hat{v} = \frac{1-r-c(1-\delta)}{(1-r)(1-\beta)}.$$

当人工智能质量足够高时，$\hat{v} < v < \bar{v}$，更高的 v 和更低的成本 δc 带来的收益超过了被贴上标签所导致的预期折价，因此所有创作者都采用人工智能。在图1中，临界值降至 $f_0 = 0$。

我们用下标 N 表示不披露制度，并用 $N_N(i)$表示选择 $i \in \{NAI,\ AI\}$的创作者比例。那么，当 $\underline{v}_1 < v < \hat{v}$时，

$$N_N(NAI) = f_0, \qquad N_N(AI) = 1 - f_0,$$

而当$\hat{v} < v < \bar{v}$时，$N_N(AI) = 1$。

平台利润与比较静态分析。对创作者的佣金收入进行汇总可得 s:

$$\Pi_N = \begin{cases} \dfrac{r\left((1-r)^2\left(2\beta v + (1-v)^2\right) - c^2(1-\delta)^2\right)}{2\beta(1-r)^2 v}, & \underline{v}_1 < v < \hat{v}, \\ \dfrac{(2-\beta)rv}{2}, & \hat{v} < v < \bar{v}. \end{cases} \tag{3}$$

命题1 随后刻画平台利润如何随人工智能价值 v 和检测准确率 β变化 .

命题1。 在不披露制度下

(i) 当 $\underline{v}_1 < v < \min\{\tilde{v}_1,\ \hat{v}\}$ 时，平台利润随 v 增加而下降；当 $\min\{\tilde{v}_1, \hat{v}\} < v < \bar{v}$时，平台利润随 v 增加而上升。
(ii) 当 $\underline{v}_1 < v < \min\{\tilde{v}_2, \hat{v}\}$ 时，平台利润随 β 增加而上升；当 $\min\{\tilde{v}_2, \hat{v}\} < v < \bar{v}$时，平台利润随 β 增加而下降。

阈值 $\tilde{v}_1$ 和 $\tilde{v}_2$ 见附录。

其经济直觉与图 1中的分类一致。当 v 处于中间范围时，v 的边际增加会吸引更多创作者采用人工智能（ f_0更小），但取代人类创作的边际人工智能内容仍很可能被折价，从而降低实际互动。由于平台不会将创作者节省的成本纳入考量，平台利润可能下降，即使创作者采用人工智能的意愿越来越强。一旦 v 足够大，即使人工智能内容被标记并折价，其仍具有价值；而更高的 v 会增加平台利润。

检测准确率 β 的影响同样是非单调的。当 AI 内容价值不高时，较高的 β 会抑制低-f 创作者的采用，并提升内容池的平均参与度，从而增加利润。当 AI 内容价值较高时，较高的 β 会增加标注和折价的发生率，降低 AI 内容的实际价值，从而减少利润。

4.2 强制披露制度

现在，我们分析平台强制披露 AI 使用时的均衡。在这一机制下，创作者可在 NAI（非 AI）、$AINAI$（使用 AI 但隐瞒）和 $AIAI$（使用 AI 并披露）之间进行选择，具体如效用（2）所示。创作者的选择取决于 AI 内容价值 v 以及在未披露后检测到 AI 使用时所施加的处罚 p。

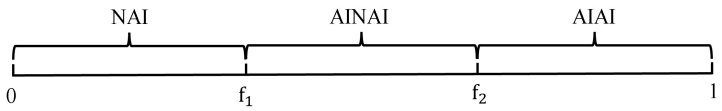

图2：披露情形下创作者的细分

创作者在 f（图2）中的细分。对于给定的 $(v,\ p)$，均衡表现为基于可信度参数 $f \in [0,1]$ 的临界点排序。图2 展示了所有策略均存在时典型的三段式结构：低可信度f创作者选择 NAI，中等可信度f创作者选择隐瞒（$AINAI$），高可信度f创作者选择披露（$AIAI$）。相关临界点为

$$f_1 = \frac{(1-r)\big(1-v(1-\beta)+p\beta\big)-c(1-\delta)}{(1-r)kv\beta}, \qquad f_2 = \frac{v-(p+v)\beta}{v-kv\beta}, \qquad f_{12} = \frac{1-r-c(1-\delta)}{(1-r)v},$$

其中，f_1是 NAI与 $AINAI$之间的无差异点，f_2是 $AINAI$与 $AIAI$之间的无差异点，而 f_{12}是 NAI与 $AIAI$之间的无差异点。

我们还定义

$$\underline{v}_2 = \frac{(1-k\beta)\big(1-r-c(1-\delta)\big)}{(1-\beta)(1-r)}, \qquad \hat{p} = \frac{c(1-\delta)+\big((1-\beta)v-1\big)(1-r)}{\beta(1-r)},$$

$$\overline{p} = \frac{c(1-\delta)(1-\beta k)+(1-r)(\beta k-\beta v+v-1)}{\beta r}.$$

给定 p 时的平台利润。根据均衡细分，平台利润等于按实际互动计算的佣金收入。我们用下标 D 表示自我披露制度。根据诱导出的策略组合，平台利润可写为如下形式。

情形 1： $\underline{v}_1 < v \le \underline{v}_2$,

$$\Pi_D(p) = r\left(\int_0^{f_{12}} 1\,df + \int_{f_{12}}^1 f v\,df\right). \tag{4}$$

情形 2： $\underline{v}_2 < v < \hat{v}$,

$$\Pi_D(p) = \begin{cases} r\left(\int_0^{f_1} 1\,df + \int_{f_1}^{f_2} \left(\beta f k v + (1-\beta)v\right) df + \int_{f_2}^{1} f v\,df\right), & 0 \le p < \overline{p}, \\ r\left(\int_0^{f_{12}} 1\,df + \int_{f_{12}}^{1} f v\,df\right), & p \ge \overline{p}. \end{cases} \tag{5}$$

情形 3： $v \ge \hat{v}$,

$$\Pi_D(p) = \begin{cases} r\left(\int_0^{f_2} \left(\beta f k v + (1-\beta)v\right) df + \int_{f_2}^{1} f v\,df\right), & 0 \le p \le \hat{p}, \\ r\left(\int_0^{f_1} 1\,df + \int_{f_1}^{f_2} \left(\beta f k v + (1-\beta)v\right) df + \int_{f_2}^{1} f v\,df\right), & \hat{p} < p < \overline{p}, \\ r\left(\int_0^{f_{12}} 1\,df + \int_{f_{12}}^{1} f v\,df\right), & p \ge \overline{p}. \end{cases} \tag{6}$$

平台选择 p 以最大化 $\Pi_D(p)$ 在 (4)–(6) 中。下一个命题刻画了平台的最优处罚。

命题 2。 在强制披露制度下，平台的均衡最优处罚为 i s

$$p^* = \begin{cases} \dfrac{c(1-\delta)(1-\beta k) + (1-r)(\beta k - 1 + (1-\beta)v)}{\beta(1-r)}, & \underline{v}_2 < v < \min\{\tilde{v}_3, \tilde{v}_4\}, \\ \dfrac{c(1-\delta)(1-\beta k)}{\beta(1-r)}, & \tilde{v}_3 < v < \tilde{v}_5, \\ 0, & \tilde{v}_4 < v < \tilde{v}_3 \ \ or \ \ \max\{\tilde{v}_3, \tilde{v}_5\} < v < \hat{v}. \end{cases} \tag{7}$$

其中， $\tilde{v}_3$、 $\tilde{v}_4$ 和 $\tilde{v}_5$ 在附录中定义。

命题 2 表明，最优执法力度会随图 3所示的均衡区域而变化。当 v 相对较低时（图 3中的区域 I），均衡中仅存在 NAI和 $AIAI$：不会出现隐瞒（图 2 收缩为两个区段），执法的主要作用是通过阻止新生的 $AINAI$区段出现来维持透明度。因此，平台设定较高的处罚，以遏制隐瞒。

当 v 处于中等水平时（图3中的区域 II），隐瞒对一部分创作者来说变得更有吸引力，从而形成图2所示的三方划分。在此区域，若要彻底消除 $AINAI$，就需要施加过高的惩罚，这也会抑制有价值的 AI 采用。因此，平台选择中等惩罚，在筛查低可信度创作者与留住平台上高价值的 AI 内容之间权衡。

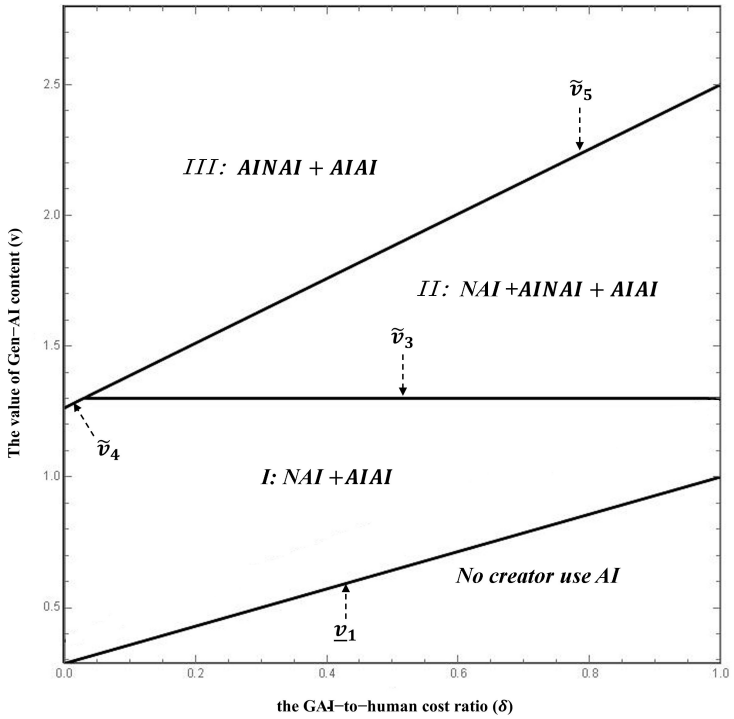

图 3：强制披露制度下，当$k = 0.8, \beta = 0.6, r = 0.3, c = 0.5$ 时创作者的均衡策略

当 v 足够高时（图3中的区域 III），人类创作会被挤出，均衡中只剩下 $AINAI$和 $AIAI$。在此区域，惩罚会适得其反，因为检测会引发标签导致的折价和信任损失，从而降低 AI 内容的实际价值。平台的最优选择是设定 $p^* = 0$，并允许创作者在隐瞒与披露之间自主选择。

推论1。 在强制披露制度下，均衡策略集随图 *3* 中的区域变化：

1. 如果 $v_1 < v < \tilde{v}_3$（区域 *I*），创作者选择 $\{NAI, AIAI\}$。

2. 如果 $\tilde{v}_3 < v < \tilde{v}_5$（区域 *II*），创作者选择 $\{NAI, AINAI, AIAI\}$。

3. 如果 $v > \tilde{v}_5$（区域 *III*），创作者选择 $\{AINAI, AIAI\}$。

推论 1 将图 3中的演变过程形式化。随着人工智能变得更有价值（和/或更具成本效益，因为 δ 下降），均衡会从透明采用阶段（区域 I）转变为存在策略性隐瞒的混合阶段（区域 II），最终进入人类创作消失的人工智能主导阶段（区域 III）。图 2 展示了相应的区域内排序情况，其中 f。

将 p^* 代入（4）–（6），可得强制披露下平台的均衡利润：

$$\Pi_D^* = \begin{cases} \frac{r\left((1-r)^2\left(v^2+1\right)-c^2(1-\delta)^2\right)}{2(1-r)^2 v} & \text{if } \underline{v_1} < v < \min\{\tilde{v}_3, \tilde{v}_4\} \\ \frac{r((1-r)^2\left(v^2\left(1+\beta\left(\beta-\beta k^2+k-2\right)\right)+(1-\beta k)(1-2(1-\beta)v)\right)-\beta c^2 k(1-\delta)^2(1-\beta k))}{2\beta k(1-r)^2(1-\beta k)v} & \text{if } \tilde{v}_3 < v < \tilde{v}_5 \\ \frac{rv(2-(2-\beta+k)\beta)}{2(1-\beta k)} & \text{if } \tilde{v}_4 < v < \tilde{v}_3 || \max\{\tilde{v}_3, \tilde{v_5}\} < v < \overline{v}. \end{cases} \tag{8}$$

在刻画均衡执行之后，我们接下来研究 AI 内容价值 v 和检测准确率 β 对 Π_D^*的影响。

命题 3. 在强制自我披露政策下，平台的均衡利润呈现以下比较静态分析：

1. ***AI*** 内容价值（v）。如果 *AI* 带来显著的成本节约（$0 < \delta < \delta_0$），那么当 $v_1 < v < v_D$ 时，平台利润随 v 增加而下降；当 $v_D < v < \overline{v}$时，平台利润随 v 增加而上升。如果成本节约有限（$\delta_0 < \delta < 1$），那么当 $\overline{v}_1 < v < v_D$时，平台利润随 v 增加而下降；当 $v_D < v < \tilde{v}_3$时，平台利润上升；当 $\tilde{v}_3 < v < \min\{v_8, \tilde{v}_5\}$时，平台利润再次下降；最后，当 $\min\{v_8, \tilde{v}_5\} < v < \overline{v}$时，平台利润上升。

2. 检测准确率（β）。当 $v_1 < v < \min\{\tilde{v}_3, \tilde{v}_4\}$时，平台利润不受 β 影响；当 $\min\{\tilde{v}_3, \tilde{v}_4\} < v < \overline{v}$ 时，平台利润随 β 增加而严格下降。

命题3(i) 表明，在强制披露制度下，平台利润随人工智能生成内容价值的变化呈非单调关系。当人工智能质量起初较低时，v 的上升会促使更多边际创作者采用人工智能，但由于披露引发的可信度折扣和信任惩罚，这些创作者的内容实现价值仍相对较低。因此，平台内容的平均质量下降，尽管技术有所改进，平台利润仍会下滑。

随着 v 持续上升，人工智能生成内容的内在质量最终会超过这些可信度损失，平台利润也开始增加。当人工智能带来的成本节约显著时（$\delta < \delta_0$），这一转变会形成一个单一转折点，使利润与 v之间呈 U 形关系，与命题1(i) 的直觉相呼应。相较之下，当人工智能带来的成本节约有限时（$\delta > \delta_0$），平台会在一段 v范围内最优地采用中等惩罚。这种内生调整会导致利润进一步下滑：随着人工智能能力提升，越来越多中等可信度创作者采取策略性不披露（*AINAI*），使内容质量下降；直到惩罚完全放宽后，利润才最终回升。由此形成 W 形走势。

命题 3(ii) 刻画了检测准确率的作用。当人工智能生成内容的价值较低时，均衡创作者策略不受检测精度影响，因此平台利润对 β 保持不变。

然而，当AI内容的价值足够高时，检测准确率提高会严格降低平台利润。检测能力提升会增加高信誉创作者在隐瞒AI使用时被标记的可能性，进而使其内容受到信誉与信任折扣，降低其实际价值。因此，更严格的检测会限制平台上高质量AI内容的供给，从而降低整体盈利能力。

4.3 平台的最优制度选择

接下来，我们比较不同制度下的均衡利润，并刻画平台的最优披露策略。命题4 总结了$(\delta,\ v)$ 空间中的均衡制度选择，其中仅在中间区域，披露制度才是最优选择。

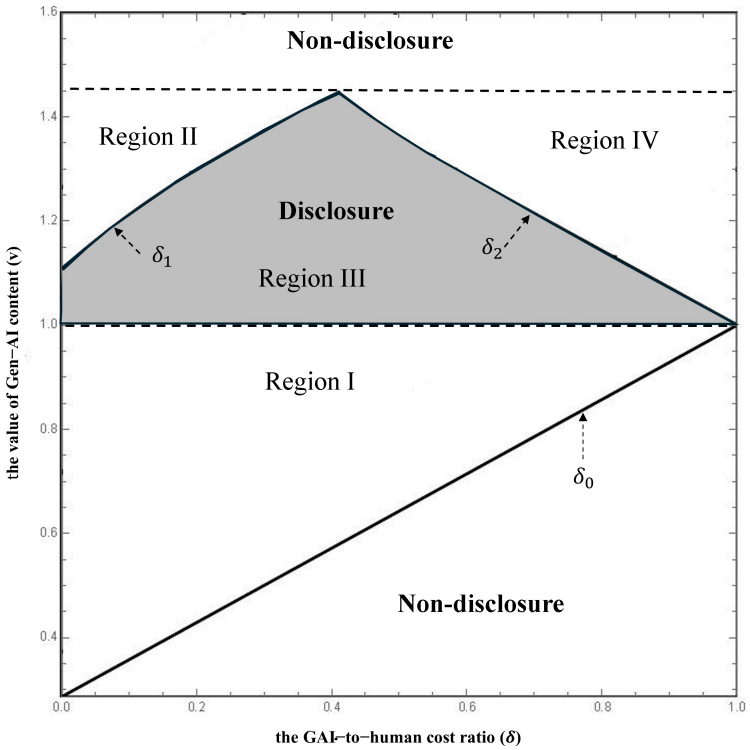

图4：当 $k=0。8$，$\beta=0。6$，$r=0。3$，$c=0。5$时平台的最优披露决策

命题4（平台的最优披露策略）。当*AI*内容价值和*AI*成本效率均处于中等水平时（即$\max\{\delta_1,0\}<\delta<\min\{\delta_0,\delta_2\}$），平台采用强制披露。否则，平台更倾向于不披露制度。

图 4 给出了相应的区域解读。在 **区域 I**（低 v）中，平台倾向于不披露，因为人工智能内容的价值不足，披露主要会抑制采用

，却无法带来相称的参与度提升。在 **区域 II**（v 非常高）中，平台再次倾向于不披露，因为披露和执行会不必要地限制高价值的 AI采用，并加剧标签导致的折价。在 **区域 III**（阴影楔形区域）中，披露是最优选择：它可充当筛选工具，限制真实性敏感型创作者采用，并减少实际价值较低的 AI内容所占比例，从而提升总体互动度。在**区域 IV**（δ 较高、成本节约有限）中，不披露更具优势，因为尽管该技术带来的成本优势相对有限，披露仍会进一步抑制采用。

4.4 对创作者、透明度和内容质量的影响

接下来，我们考察平台的披露制度如何影响创作者、信息透明度和总体内容质量。创作者利润尤其值得关注，因为创作者构成了平台的主要供给侧。创作者的参与、努力程度以及对内容质量的长期投资，都在很大程度上取决于预期回报。因此，系统性改变创作者剩余的披露政策，不仅可能影响短期结果，还可能影响内容生态系统的可持续性。

创作者利润。令 CS_i 表示制度 $i \in \{N, D\}$下的创作者总利润。利用均衡策略，

$$
\begin{aligned}
CS_N &= \int_0^{f_0} \big((1-r)-c\big)\,df + \int_{f_0}^{1} \Big((1-r)\big(\beta f v + (1-\beta)v\big) - \delta c\Big)\,df, \\
CS_D &= \int_0^{f_1} \big((1-r)-c\big)\,df + \int_{f_1}^{f_2} \Big((1-r)\big(\beta(fkv-p) + (1-\beta)v\big) - \delta c\Big)\,df + \int_{f_2}^{1} \Big((1-r)fv - \delta c\Big)\,df.
\end{aligned}
\tag{9}
$$

基于方程 9，我们分析自我披露对创作者利润的影响，具体见命题 5。

命题 5。平台的披露政策会降低创作者总利润。（即，$CS_D < CS_N$）。

尽管自我披露通常被视为一种提升透明度的政策，但我们的分析（命题 4 和 5）表明，它也发挥着战略性治理工具的作用。平台通过对创作者施加披露相关成本，在生态系统层面将信任外部性内部化，从而提高自身收益，即使创作者总剩余有所下降。

透明度与总体内容质量。 根据Zhu(2002年)，我们将透明度定义为 1 减去 AI内容未被标注而出现的概率：

$$\begin{aligned} T_N &= 1-(1-\beta)\,N_N(AI), \\ T_D &= 1-(1-\beta)\,N_D(AINAI). \end{aligned} \tag{10}$$

披露通过减少逃过检测的未披露AI内容，从而提高透明度（$T_D > T_N$）。

接下来，我们将总体内容质量定义为所有创作者的质量贡献之和：

$$\begin{aligned} Q_N &= N_N(NAI) + v\,N_N(AI), \\ Q_D &= N_D(NAI) + v\,N_D(AINAI) + v\,N_D(AIAI). \end{aligned} \tag{11}$$

命题6。 平台的披露政策会对透明度和总体内容质量产生如下影响：

1. 披露会严格提高透明度（即 $T_D > T_N$）。

2. 当人工智能生成内容的技术水平低于人类创作时（即 $Q_D > Q_N for\, v_1 < v < \min\{\tilde{v}_4, 1\}$），披露当且仅当满足该条件时才会提高总体内容质量；而当人工智能生成内容的技术水平更高时（即 $Q_D < Q_N for\, (1 < v < \tilde{v}_4 || 1 < v < \tilde{v}_5)$），披露会降低总体内容质量。否则，$Q_D = Q_N$。

质量方面的结果体现了构成效应，这是因为披露改变了创作者之间生产技术的均衡分配。披露提高了人工智能使用的预期成本，因而抑制边际创作者采用人工智能，并使生产转向人类创作或实际互动较低的标注AI内容。因此，总体内容质量会随着内容生产者及其生产方式的权重变化而改变。当人工智能质量较低时（$v < 1$），这种重新分配会挤出质量较差的人工智能生成内容，从而改善内容池。当人工智能质量较高时（$v > 1$），同一机制会抑制质量较高的人工智能产出，尽管透明度有所提高，总体内容质量仍会下降。

这些结果表明，基于披露的治理需要在透明度与生产效率之间权衡。自我披露通过限制未披露的 AI 使用，能够可靠地降低信息不透明性，但也提高了 AI 采用的实际成本。当人工智能质量较低时，这种扭曲是有益的，因为它会挤出劣质的 AI 内容，并提升总体质量。当人工智能质量较高时，同样的扭曲会抑制优质 AI 内容的生产，导致内容质量下降，尽管透明度有所提升。只有

当人工智能质量和成本效率处于中间水平时，披露才能发挥有效筛选工具的作用，在提升透明度的同时维持生产效率。

5 结论

基于均衡分析，我们就数字平台上的人工智能生成内容治理得出若干结论。核心观点是，披露和执行政策塑造结果的方式，不仅在于改善信息，还在于重新配置创作者之间的生产技术和剩余；而其影响在很大程度上取决于人工智能的技术成熟度。

首先，尽管随着人工智能质量的提升，人工生成内容最终会被挤出，但这一结果的实现路径取决于人工智能工具的成本效率。当人工智能能够带来显著的成本节约时，平台会最优地突然放弃执行，通过取消惩罚，充分利用广泛采用人工智能所带来的规模收益和生产率收益。当成本效率较为有限时，平台则会逐步放松执行。在这种情况下，惩罚充当一种临时性的筛选工具，在允许逐步提升的人工智能质量得到吸收的同时，抑制低可信度人工智能采用。这一中间制度通过在短期信任维护与长期效率收益之间取得平衡，使向人工智能主导的过渡更加平稳。

其次，只有当人工智能生成内容的内在价值和其节省成本优势均处于中间区间时，平台才会采用自我披露制度。在这一范围内，披露作为一种有效的治理工具发挥作用：它提高透明度，将人工智能使用重新配置给内容较不易受到可信度担忧影响的创作者，并维持总体内容质量。超出这一范围后，披露会适得其反。当人工智能质量较低时，披露主要抑制采用，却无法带来有意义的质量改善。当人工智能质量较高时，披露通过对优质人工智能内容施加估值折扣，扭曲高效采用；尽管透明度提高，却降低了总体内容质量和创作者剩余。

最后，我们的分析表明，尽管自我披露旨在提升透明度，但它在不同市场参与者之间并非中性。平台会策略性地运用披露来管理信任外部性和内容构成，并经常将合规成本与可信度成本转嫁给创作者。因此，披露可能在提高平台利润的同时，减少创作者剩余，并在某些地区抑制高质量AI内容。这种张力表明，透明度的提升可能以牺牲创作者激励和总体内容质量为代价，从而引发人们对披露驱动的治理能否长期持续的担忧。

参考文献

Bhargava HK（2022）创作者经济：管理生态系统供给、收益分成与平台设计。*Management Science* 68（7）：5233–5251。Brüns JD、Meißner M（2024）你是自己创作内容的吗？使用生成式人工智能进行社交媒体内容创作会降低感知到的品牌真实性。《零售与消费者服务期刊》79：103790。Bubeck S、Chadrasekaran V、Eldan R、Gehrke J、Horvitz E、Kamar E、Lee P、Lee YT、Li Y、Lundberg S等（2023）通用人工智能的曙光：GPT-4的早期实验。*arXiv*预印本 *arXiv:2303.12712*。Bunescu RC、Uduehi OO（2019）学会制造惊喜：作曲者—受众架构。国际计算创造力会议论文集，41–48。Castelo N、Bos MW、Lehmann DR（2019）任务依赖型算法厌恶。《营销研究期刊》56（5）：809–825。Cavusoglu H、Raghunathan S、Cavusoglu H（2009）信息安全技术的配置及其交互：防火墙与入侵检测系统案例。《信息系统研究》20（2）：198–217。Chen H、Hu YJ、Huang S（2019）社交媒体上的金钱激励与股票观点。《管理信息系统期刊》36（2）：391–417。Chen Z、Chan J（2024）创造性工作中的大型语言模型：协作模式与用户专业水平的作用。*Management Science* 70（12）：9101–9117。Dietvorst BJ、Simmons JP、Massey C（2015）算法厌恶：人们在看到算法出错后会错误地回避算法。《实验心理学期刊：总论》144（1）：114–126。Eloundou T、Manning S、Mishkin P、Rock D（2023）GPT就是GPT：初探大型语言模型对劳动力市场的潜在影响。*arXiv*预印本 *arXiv:2303.10130* 。Fang X、Che S、Mao M、Zhang H、Zhao M、Zhao X（2024）人工智能生成内容的偏见：对大型语言模型生成新闻的考察。《科学报告》14（1）：5224。Fügener A、Grahl J、Gupta A、Ketter W（2022）人类与人工智能协作中的认知挑战：探究通往有效委派的路径。《信息系统研究》33（2）：678–696。Han Y、Lappas T、Sabnis G（2020）研究社交媒体传播性时内容特征与创作者特征交互的重要性。《信息系统研究》31（2）：576–588。Jiménez-Durán R（2023）内容审核的经济学：来自Twitter仇恨言论的理论与实验证据。工作论文，芝加哥大学，美国伊利诺伊州芝加哥。Kalpesh K、Yixiao S、Marzena K、John W、Mohit I（2023）释义改写能够规避人工智能生成文本检测器，但检索是一种有效的防御手段。*arXiv*预印本 *arXiv:2303.13408* 。

Kim A, Dennis AR (2019) 谁在发声？呈现格式和来源评级对社交媒体假新闻的影响。*MIS Quarterly* 43(3):1025–1040. Kirchenbauer J, Geiping J, Wen Y, Katz J, Miers I, Goldstein T (2023) 大型语言模型的水印。国际机器学习大会, 17061–17084. Liu M, Tang X, Xia S, Zhang S, Zhu Y, Meng Q (2023) 算法厌恶：来自网约车司机的证据。*Management Science* 72(1):193–203. Liu Y, Yildirim P, Zhang ZJ (2022) 收入模式和技术对内容审核策略的影响。*MarketingScience* 41(4):831–847. Logg JM, Minson JA, Moore DA (2019) 算法欣赏：人们更偏好算法判断，而非人类判断。*Organizational Behavior and Human Decision Processes*151:90–103. Longoni C, Bonezzi A, Morewedge CK (2019) 对医疗人工智能的抵制。*Journal of Consumer Research* 46(4):629–650. Luo X, Tong S, Fang Z, Qu Z (2019) 前沿：机器与人类之争：人工智能聊天机器人披露对顾客购买的影响。*Marketing Science*38(6):937–947. Madio L, Quinn M (2024) 社交媒体平台上的内容审核与广告。*Journal of Economics &Management Strategy* 34(2):342–369. Noy S, Zhang W (2023) 生成式人工智能生产力效应的实验性证据。*Science*381(6654):187–192. Ouyang L, Wu J, Jiang X, Almeida D, Wainwright C, Mishkin P, Zhang C, Agarwal S, Slama K, Ray A, et al. (2022) 训练语言模型遵循人类反馈给出的指令。*Advances inNeuralInformationIrocessingSystems* 35:27730–27744. Papanastasiou, Yiangos (2020) 假新闻传播与检测：一个序贯模型。*Management Science*66(5):1826–1846. Pérez-Rosas V, Kleinberg B, Lefevre A, Mihalcea R (2017) 自动检测假新闻。*arXiv preprintarXiv:1708.07104* . Sadasivan VS, Kumar A, Balasubramanian S, Wang W, Feizi S (2023) 能否可靠地检测人工智能生成文本？*arXivpreprintarXiv:2303.11156* . Schanke S, Burtch G, Ray G (2024) 数字拟声鸟：语音深度伪造影响信任的实验性证据。*Management Science* 72(1):386–405. Sietsema J, Ghoseyni K, Condron C, Saks K, Spradling M (2023) 一项关于用户对新闻媒体标注的感知的研究：以限制错误信息传播。*2023*年国际计算科学与计算智能大会, 719–726. Tiwari AK, Sharma A, Rayakar P, Bhavriya MK, et al. (2024) 人工智能生成视频的伪造检测与认证。*2024*年*IEEE*第九届技术融合国际会议, 1–8.

Turel O, Kalhan S (2023) 对机器抱有偏见？隐性联想与算法厌恶的短暂性。*MIS Quarterly*47(4):1369–1394. Wan J, Lu Y, Wang B, Zhao L (2017) 依恋如何影响用户在社交媒体上向内容创作者捐赠的意愿：社会技术系统视角。《信息与管理》 54(7):837–850. Wang M, Jung J, Bapna R (2025) 真实感与人工智能披露对虚拟网红效果的影响：一项大规模实地实验。工作论文，天普大学福克斯商学院，费城，宾夕法尼亚州。Wang SA, Min-Seok, Pavlou, A P (2022) 眼见为实？在假新闻中加入视频如何影响用户向社交媒体平台举报假新闻。*MIS Quarterly* 46(3):1323–1354.

D. Weber-Wulff，A. Anohina-Naumeca，S. Bjelobaba，Folt`ynek T，J. Guerrero-Dib，O. Popoola，P. Šigut，L. Waddington（2023）人工智能生成文本检测工具测试。《国际教育诚信期刊》19(1)：1–39。

F. Yao，C. Li，D. Nekipelov，H. Wang，H. Xu（2024）内容创作竞争中的人类与生成式人工智能：共生还是冲突？*arXiv*预印本 *arXiv:2402.15467* 。

J. H. Ye，C. E. H. Chua（2022）社交媒体平台上内容生成和用户参与度的变现：来自付费问答的证据。*IEEE*工程管理汇刊 71：4022–4034。

Y. Yin，N. Jia，C. J. Wakslak（2024）人工智能可以帮助人们感到自己被倾听，但人工智能标签会削弱这一效果。《美国国家科学院院刊》121(14)：e2319112121。

K. Zhu（2002年）信息透明度在电子市场中的作用：为什么数据透明度可能阻碍企业间交易平台的采用。《电子市场》12(2):92–99。

附录：证明

命题 1的证明

当 $0 < v < v_1$、$\frac{\partial \Pi_D}{\partial v} = 0$ 时，利润与 v 无关。当 $v_1 < v < \hat{v}$、$\frac{\partial \Pi_D}{\partial v} = r(c^2(1-\delta)^2+(1-r)^2(v^2-1))\,2\beta(1-r)^2v^2$。我们可以得出，$\frac{\partial \Pi_D}{\partial v} < 0$ 当且仅当 $v_1 < v < \min\{\tilde{v}_1, \bar{v}\}$，且 $\frac{\partial \Pi_D}{\partial v} > 0$ 当且仅当 $\tilde{v}_1 < v < \hat{v}$。当 $v > \hat{v}$、$\frac{\partial \Pi_D}{\partial v} = (2-\beta)_2^{\,r} > 0$ 时。其中 $\tilde{v}_1 = \frac{\sqrt{(1-r)^2-c^2(1-\delta)^2}}{1-r}$。因此，平台关于 v 的利润函数先与 v 无关，随后递减，最后递增。

当 $0 < v < v_1$、$\frac{\partial \Pi_D}{\partial \beta} = 0$ 时，利润与 β 无关。当 $v_1 < v < \hat{v}$ 时，$\frac{\partial \Pi_D}{\partial \beta} = r(c^2(1-\delta)^2-(1-r)^2(v-1)^2)\,2\beta^2(1-r)^2v$。我们可以得出结论：当且仅当 $v_1 < v < \min\{\hat{v}, \tilde{v}_2\}$ 时，$\frac{\partial \Pi_D}{\partial \beta} > 0$；当且仅当 $\tilde{v}_2 < v < \hat{v}$ 时，$\frac{\partial \Pi_D}{\partial v} < 0$。其中 $\tilde{v}_2 = \frac{1-r+c(1-\delta)}{1-r}$。当 $v > \hat{v}$ 时，$\frac{\partial \Pi_D}{\partial \beta} = \frac{-rv}{2} < 0$。因此，平台的利润函数关于 v 先与 v 无关，随后递增，最后递减。

■

命题 2的证明

平台在自我披露政策下实现利润最大化。根据 v 的大小，我们将分析分为三种情况：

情形 1：当 $v_1 < v < v_2$，根据方程 5，没有创作者采用 $AINAI$ 策略，因此平台无需施加任何处罚。

情形 2：当 $v_2 \le v < \hat{v}$，根据方程 6：

1. $0 < p < \bar{p}$，平台的利润关于 p 是一个凹函数：

$$\pi_D = -\frac{\beta r p^2}{2kv(1-\beta k)} + \frac{c(1-\delta)rp}{k(1-r)v} + \frac{(r-1)^2v^2\left(\beta\left(\beta\left(k^2-1\right)-k+2\right)-1\right)}{2\beta k(1-r)^2v(\beta k-1)} + \frac{r\left(c^2(1-\delta)^2(1-\beta k)+\beta k(1-r)^2+2(\beta-1)(1-r)^2v(\beta k-1)-(1-r)^2\right)}{2\beta k(r-1)^2v(\beta k-1)}$$

我们可以重新整理 $\left.\frac{\partial \pi_D}{\partial p}\right|_{p=0} = \frac{c(1-\delta)k}{1-r}(v) > 0$，但 $\left.\frac{\partial \pi_D}{\partial p}\right|_{p=\bar{p}} = \frac{r(1-v-\beta k+\beta v)}{kv(1-\beta k)}$，该表达式的符号无法确定。当 $\frac{1-\beta k}{1-\beta} < v < \frac{1-r-c}{1-\beta}(1-\delta)$ 时，$\left.\frac{\partial \pi_D}{\partial p}\right|_{p=\bar{p}} < 0$，且 π_D 随 p 先增后减。因此，最优处罚 $p^* = \frac{c(1-\delta)}{(1-\beta k)\beta(1-r)}$，其在 $\frac{\partial \pi_D}{\partial p} = 0$ 处取得。当 $v_2 < v \le \min\{\frac{1-\beta k}{1-\beta}, \frac{1-r-c+c\delta}{1-r-\beta+\beta r}\}$ 时，$\left.\frac{\partial \pi_D}{\partial p}\right|_{p=\bar{p}} > 0$，且 π_D 随 v 增加。因此，最优处罚 $p^* = \bar{p}$。

2. $p > \bar{p}$, $\Pi_D = \frac{r\left((1-r)^2(v^2+1)-c^2(1-\delta)^2\right)}{2(1-r)^2v}$，且整个函数连续。

因此，情形 2 的结果为：

$$p_1^* = \begin{cases} \frac{c(1-\delta)(1-\beta k)+(1-r)(\beta k-1+(1-\beta)v)}{\beta(1-r)} & v_2 < v \le \min\{\tilde{v}_3, \hat{v}\} \\ \frac{c(1-\delta)(1-\beta k)}{\beta(1-r)} & \tilde{v}_3 < v < \hat{v} \end{cases} \tag{12}$$

情形 3：当 $\hat{v} < v < \bar{v}$ 且根据方程7：

1. 当 $0 < p < \hat{p}$ 时，平台关于 p 的利润为凹函数：$\Pi_D = \frac{r\left(-\beta^2p^2+v^2(2-\beta(2-\beta+k))\right)}{2v(1-\beta k)}$，且 $\left.\frac{\partial \pi_D}{\partial p}\right|_{p=\hat{p}} = 0$，因此当 $0 < p < \hat{p}$ 时，Π_D 随 p 增加而减少，且 $p^* = 0$。

2. 当 $\hat{p} < p < \bar{p}$ 时，

$$\pi_D = -\frac{\beta r p^2}{2kv(1-\beta k)} + \frac{c(1-\delta)rp}{k(1-r)v} + \frac{(r-1)^2v^2\left(\beta\left(\beta\left(k^2-1\right)-k+2\right)-1\right)}{2\beta k(1-r)^2v(\beta k-1)} + \frac{r\left(c^2(1-\delta)^2(1-\beta k)+\beta k(1-r)^2+2(\beta-1)(1-r)^2v(\beta k-1)-(1-r)^2\right)}{2\beta k(r-1)^2v(\beta k-1)}$$

首先，$\frac{\partial \pi_D}{\partial p}\big|_{p=\hat{p}} = \frac{r}{(1-r-v-\beta ck+\beta(c\delta k+v)+(1-\beta)rv)\, k(1-r)v(1-\beta k)}$，在 $\hat{v} < v < v$ 的限制下，我们可以得出，当 $\hat{v} < v < 1-r-\beta ck(1-\delta)\,(1-r)\,(1-\beta)$ 时，$\frac{\partial \pi_D}{\partial p}\big|_{p=\hat{p}} > 0$；而当 $1-r-\beta ck(1-\delta)\,(1-r)\,(1-\beta) < v < v$ 时，$\frac{\partial \pi_D}{\partial p}\big|_{p=\hat{p}} < 0$。其次，$\frac{\partial \pi_D}{\partial p}\big|_{p=p} = r(1-v-\beta k+\beta v)\, kv(1-\beta k)$。我们可以看出，在 $\hat{v} < \bar{v} < v$ 的限制下，$\frac{\partial \pi_D}{\partial p}\big|_{p=p} > 0$ 当且仅当 $\hat{v} < v < \tilde{v}_3$，且 $\frac{\partial \pi_D}{\partial p}\big|_{p=p} < 0$ 当且仅当 $\min\{\hat{v}, \tilde{v}_3\} < v < v$。因此，当 $\hat{v} < v < \tilde{v}_3$ 时，π_D 随 p 增大，且 $p^* = p$。当 $\max\{\hat{v}, \frac{1-k\beta}{1-\beta}\} < v < 1-r-\beta ck(1-\delta)\,(1-r)\,(1-\beta)$ 时，函数 π_D 随 p 先增后减，且 $p^* = c(1-\delta)\,(1-\beta k)\,\beta(1-r)$，其在 $\frac{\partial \pi_D}{\partial p} = 0$ 处取得。当 $1-r-\beta ck(1-\delta)\,(1-r)\,(1-\beta) < v < v$ 时，π_D 随 p 减小，$p^* = \hat{p}$。

3. 当 $p > p, \Pi_D = r(\frac{(1-r)^2(v^2+1)-c^2(1-\delta)^2}{2(1-r)^2 v}$ 时，且整个函数连续。

因为当 $0 < p < \hat{p}$ 时，Π_D 随着 p 的增加而减小，且 Π_D 在整个范围内连续。因此，$\hat{p}$ 不会是最优决策。首先，当 $1-r-\beta ck(1-\delta)\,(1-r)\,(1-\beta) \le v < v$ 时，$p^* = 0$。其次，当 $\hat{v} < v < \tilde{v}_3$ 时，我们需要比较 $\pi_D\big|_{p=0} = \frac{rv(2-\beta(2-\beta+k))}{2-2\beta k}$ 和 $\pi_D\big|_{p=p} = \frac{r((1-r)^2(1+v^2)-c^2(1-\delta)^2)}{2(1-r)^2 v}$。我们可以得出结论：当 $\hat{v} < v < \min\{\tilde{v}_3, \tilde{v}_4\}$ 时，$\Pi_D\big|_{p=p} > \Pi_D\big|_{p=0}$。当 $\tilde{v}_4 < v < \tilde{v}_3$ 时，$\Pi_D\big|_{p=p} < \Pi_D\big|_{p=0}$。第三，当 $\max\{\hat{v}, \tilde{v}_3\} \prec v < 1-r-\beta ck(1-\delta)\,(1-r)\,(1-\beta)$ 时，我们需要比较 $\pi_D\big|_{p=0} = \frac{rv(2-\beta(2-\beta+k))}{2-2\beta k}$ 和 $\pi_D\big|_{p=c(1-\delta)\,(1-\beta k)\,\beta(1-r)} = r(\beta c^2(\delta-1)^2 k(\beta k-1)+(1-r)^2(v^2(\beta(\beta-\beta k^2+k-2)+1)-\beta k-2(\beta-1)v(\beta k-1)+1))\, 2\beta k(1-r)^2 v(1-\beta k)$。我们可以得出结论：当 $\max\{\hat{v}, \tilde{v}_3\} < v < \tilde{v}_5$ 时，$\Pi_D\big|_{p=c(1-\delta)\,(1-\beta k)\,\beta(1-r)} > \Pi_D\big|_{p=0}$；而当 $\max\{\tilde{v}_5, \tilde{v}_3\} < v < 1-r-\beta ck(1-\delta)\,(1-r)\,(1-\beta)$ 时，$\Pi_D\big|_{p=c(1-\delta)\,(1-\beta k)\,\beta(1-r)} > \Pi_D\big|_{p=0}$。

Where $\tilde{v}_3 = \frac{1-k\beta}{1-\beta}$, $\tilde{v}_4 = \frac{\sqrt{1-\beta k}\sqrt{(1-r)^2-c^2(1-\delta)^2}}{(1-\beta)(1-r)}$, $\tilde{v}_5 = \frac{1}{1-\beta} - \frac{c(1-\delta)\sqrt{\beta k}}{(1-\beta)(1-r)}$.

因此，我们得到情形 3 的结果：

$$p_2^* = \begin{cases} \frac{c(1-\delta)(1-\beta k)+(1-r)(\beta k-1+(1-\beta)v)}{\beta(1-r)} & \text{if } \hat{v} < v < \min\{\tilde{v}_3, \tilde{v}_4\} \\ \frac{c(1-\delta)(1-\beta k)}{\beta(1-r)} & \text{if } \max\{\hat{v}, \tilde{v}_3\} < v < \tilde{v}_5 \\ 0 & \text{if} \tilde{v}_4 < v < \tilde{v}_3 || \max\{\tilde{v}_5, \tilde{v}_3\} < v < \bar{v}. \end{cases} \tag{13}$$

令 $p = c(1-\delta)\,(1-\beta k)+(1-r)\,(\beta k-1+(1-\beta)v)\,\beta(1-r)$ 且 $\tilde{p} = c(1-\delta)\,(1-\beta k)\,\beta(1-r)$。

结合方程 12 和 13中的结果，我们可以得出平台的最优处罚决策如下：

$$p^* = \begin{cases} \frac{c(1-\delta)(1-\beta k)+(1-r)(\beta k-1+(1-\beta)v)}{\beta(1-r)} & \text{if } \underline{v}_2 < v < \min\{\tilde{v}_3, \tilde{v}_4\} \\ \frac{c(1-\delta)(1-\beta k)}{\beta(1-r)} & \text{if } \tilde{v_3} < v < \tilde{v}_5 \\ 0 & \text{if } \tilde{v}_4 < v < \tilde{v}_3 || \max\{\tilde{v}_3, \tilde{v_5}\} < v < \bar{v}. \end{cases} \tag{14}$$

■

命题 3 的证明

根据方程 8 中平台的利润，当 $v_1 < v < \min\{\tilde{v}_3, \tilde{v}_4\}$ 时，$\frac{\partial \Pi_D}{\partial v} = r(c^2(1-\delta)^2 + (1-r)^2(v^2-1))\ 2(1-r)^2v^2$。其符号取决于 $c^2(1-\delta)^2 + (1-r)^2\left(v^2-1\right)$，这是一个凸函数。我们可以得出，$\frac{\partial \Pi_D}{\partial v} < 0$ 当且仅当 $v_1 < v < \frac{\sqrt{(1-r)^2 - c^2\ 1-\delta\ ^2}}{1-r}$，且 $\frac{\partial \Pi_D}{\partial v} > 0$ 当且仅当 $\frac{\sqrt{(1-r)^2 - c^2\ 1-\delta\ ^2}}{1-r} < v < \min\{\tilde{v}_3, \tilde{v}_4\}$。$\frac{\partial \Pi_D}{\partial \beta} = 0$。

当 $\tilde{v}_3 < v < \tilde{v}_5$，$\frac{\partial \Pi_D}{\partial v} = (1-r)^2\left(1-\beta\left(2-k-\left(1-k^2\right)\beta\right)\right)v^2 - r(1-\beta k)\left(1-(2-r)r-\beta c^2(1-\delta)^2 k\right)$，这是一个凸函数，并且有一个正根 $v_8 = \frac{\sqrt{1-\beta k}\sqrt{1+(r-2)}}{\sqrt{(1-r^2\ 1-\beta\ \beta\ k^2-1\ -k+2}}$。$v_8 > \tilde{v}_3$ 总是成立。因此，$\frac{\partial \Pi_D}{\partial v} < 0$ 当且仅当 $\tilde{v}_3 < v < \min\{v_8, \tilde{v}_5\}$，且 $\frac{\partial \Pi_D}{\partial v} > 0$ 当且仅当 $v_8 < v < \tilde{v}_5$。$\frac{\partial \Pi_D}{\partial \beta} = r(\beta k - \beta v + v - 1)(v(\beta(2k-1)-1)+1-\beta k)\ 2\beta^2 k v(1-\beta k)_2 < 0$。

当 $\tilde{v}_4 < v < \tilde{v}_3 || \max\{\tilde{v}_3, \tilde{v}_5\} < v < v$，$\frac{\partial \Pi_D}{\partial v} = r(2-\beta(2-\beta+k))_{2-2\beta k} > 0$ 时。$\frac{\partial \Pi_D}{\partial \beta} = -(1-\beta)rv(2-(1+\beta)k)\ 2(1-\beta k)_2 < 0$。令 $v_D = \frac{\sqrt{(1-r)^2 - c^2\ 1-\delta\ ^2}}{1-r}$。■

命题 4 的证明

我们比较两种不同政策下的利润，如方程 3 和 8所示：

(i)当 $v_1 < v < v_2$且 $v_2 < v < \min\{\hat{v}, \tilde{v}_3\}$ 时，$\pi_N^* - \pi_D^* = \frac{(1-\beta)r((1-r)^2(v-1)^2 - c^2(1-\delta)^2)}{2\beta(1-r)^2 v}$，其符号取决于 $(1-r)^2(v-1)^2 - c^2(1-\delta)^2$，，这是关于 v 的凸函数，且有一个根 $v_0 = \frac{1-r+c}{1-r}(1-\delta)$。对于 $v_1 < v < v_2$，当且仅当 $v_1 < v < \min\{v_2, v_0\}$ 时，$\pi_N^* - \pi_D^* < 0$；当且仅当 $v_0 < v < v_2$ 时，$\pi_N^* - \pi_D^* > 0$。对于 $v_2 < v < \min\{\hat{v}, \tilde{v}_3\}$，当且仅当 $v_2 < v < \min\{\hat{v}, \tilde{v}_3, v_0\}$ 时，$\pi_N^* - \pi_D^* < 0$；当且仅当 $\max\{v_2, v_0\} < v < \min\{\hat{v}, \tilde{v}_3\}$ 时，$\pi_N^* - \pi_D^* > 0$。

(ii) 当 $\tilde{v}_3 < v < \hat{v}$ 时，

$$\Delta\Pi = \frac{(1-\beta)(1-k)(1-(1+k))\beta)v^2}{k(1-\beta k)} + \frac{2(1-\beta)(1-k)}{k}v + 1 - \frac{1}{k} + \frac{(\beta-1)c^2(\delta-1)^2}{(r-1)^2}$$

该函数的两个根分别为 $v_1 = (1-\beta k)\left(k\left(\beta + \sqrt{(1-\beta)\ (1-k)\ \frac{((1-\beta)c^2(1-\delta)^2(\beta+\beta k-1)+\beta^2(1-k)(1-r)^2)\ k}{(r-1)^2(1-\beta k)}}\right)^{-1}\right) + 1 - \beta\Big)(1-\beta)\ (1-k)\ (1-(1+k)\beta)$ 和 $v_2 = (1-\beta k)\left(k\left(\beta - \sqrt{(1-\beta)\ (1-k)\ \frac{((1-\beta)c^2(1-\delta)^2(\beta+\beta k-1)+\beta^2(1-k)(1-r)^2)\ k}{(r-1)^2(1-\beta k)}}\right)^{-1}\right) + 1 - \beta\Big)(1-\beta)\ (1-k)\ (1-(1+k)\beta)$，如果 $1-(1+k))\ \beta > 0$, $\Delta\Pi$ 为凹函数

v、$v_1 > v_2$、$v_1 > \hat{v}$ 的函数关系始终成立，因此我们可以得出，$\Delta\Pi < 0$ 当且仅当 $\tilde{v}_3 < v < \min\{\hat{v}, v_2\}$时成立，而 $\Delta\Pi > 0$ 当且仅当$\max\{\tilde{v}_3, v_2\} < v < \hat{v}$时成立。当 $1-(1+k)\beta < 0$、$v_1 < v_2$、$v_1 < \tilde{v}_3$的函数关系始终成立，因此我们可以得出，$\Delta\Pi < 0$ 当且仅当 $\tilde{v}_3 < v < \min\{\hat{v}, v_2\}$ 时成立，而 $\Delta\Pi > 0$ 当且仅当$\max\{\tilde{v}_3, v_2\} < v < \hat{v}$时成立。

(iii) 当 $\hat{v} < v < \min\{\tilde{v}_3, \tilde{v}_4\}$ 时，

$$\Delta\Pi = \frac{r\left(c^2(1-\delta)^2 - (1-r)^2\left(1-(1-\beta)v^2\right)\right)}{2(1-r)^2 v}$$

$\Delta\Pi$的符号取决于 $c^2(1-\delta)^2 - (1-r)^2(1-(1-\beta)v^2)$，这是一个仅有一个位置根 $v_5 = \frac{\sqrt{(1-r)^2 - c^2(1-\delta)^2}}{\sqrt{(1-\beta)(1-r^2)}}$ 的凸函数。此外，$v_5 < \tilde{v}_4$始终成立。

我们可以得出，$\Delta\Pi < 0$ 当且仅当 $\hat{v} < v < \min\{v_5, \tilde{v}_3\}$ 时成立；$\Delta\Pi > 0$ 当且仅当 $\min\{v_5, \hat{v}\} < v < \min\{\tilde{v}_3, \tilde{v}_4\}$ 时成立。

(iv) 当 $\max\{\hat{v}, \tilde{v}_3\} < v < \tilde{v}_5$ 时，

$$\Delta\Pi = \frac{1}{2}(2-\beta)rv - \frac{r\left(\beta c^2(\delta-1)^2 k(\beta k-1) + (1-r)^2\left(v^2\left(\beta\left(\beta-\beta k^2+k-2\right)+1\right) - \beta k - 2(\beta-1)v(\beta k-1)+1\right)\right)}{2\beta k(1-r)^2 v(1-\beta k)}$$

该函数的符号与 $f(v) = (1-\beta)(1-r)^2 r(\beta(1+k(1-\beta k)) - 1)v^2 + 2(1-\beta)(1-r)^2(1-\beta k)rv - \beta c^2(\delta-1)^2 kr(\beta k-1) + \beta kr(r-1)^2 - (r-1)^2 r$。它有两个根：$v_3 = \frac{\beta^2(-k)(r-1)^2 + \beta(k+1)(r-1)^2 - (r-1)^2 + \sqrt{(\beta-1)\beta k(r-1)^2(\beta k-1)(c^2(\delta-1)^2(\beta(k(\beta k-1)-1)+1) - \beta(k-1)(r-1)^2)}}{(\beta-1)(r-1)^2(\beta(k(\beta k-1)-1)+1)}$ 和 $v_4 = \frac{\beta^2(-k)(r-1)^2 + \beta(k+1)(r-1)^2 - (r-1)^2 - \sqrt{(\beta-1)\beta k(r-1)^2(\beta k-1)(c^2(\delta-1)^2(\beta(k(\beta k-1)-1)+1) - \beta(k-1)(r-1)^2)}}{(\beta-1)(r-1)^2(\beta(k(\beta k-1)-1)+1)}$。

当 $\beta(1+k(1-\beta k))-1 > 0$，$v_3 > v_4$是关于 v 和 $v_4 < \min\{\hat{v}, \tilde{v}_3\}$ 的凸函数，且 $v_3 < \tilde{v}_5$ 恒成立。因此，$\Delta\Pi < 0$ 当且仅当 $\max\{\hat{v}, \tilde{v}_3\} < v < v_3$，而 $\Delta\Pi > 0$ 当且仅当 $\max\{\hat{v}, \tilde{v}_3, v_3\} < v < \tilde{v}_5$。当 $\beta(1+k(1-\beta k)) - 1 < 0$，$v_3 < v_4$是关于 v 和 $v_3 < \tilde{v}_5$ 的凹函数，且 $v_4 > \tilde{v}_5$ 恒成立。因此，$\Delta\Pi < 0$ 当且仅当 $\max\{\hat{v}, \tilde{v}_3\} < v < v_3$，而 $\Delta\Pi > 0$ 当且仅当 $\max\{\hat{v}, \tilde{v}_3, v_3\} < v < \tilde{v}_5$。

(v) 当 $\tilde{v}_4 < v < \tilde{v}_3 || \max\{\tilde{v}_3, \tilde{v}_5\} < v < v, \pi_N^* - \pi_D^* = \frac{\beta r v}{(1-k)(1-\beta)(2-2\beta k)} > 0$。

此外，我们分析边界 $v = \{v_0, v_2, v_3, v_5\}$。很容易看出 $\frac{\partial v_5}{\partial\delta} > 0$，$\frac{\partial v_3}{\partial\delta} > 0$，$\frac{\partial v_2}{\partial\delta} < 0$和 $\frac{\partial v_0}{\partial\delta} < 0$。$v_3 = v_5$的根为 $\delta = 1 - (1-r)\sqrt{(1-\beta)(k(2-\beta k)-1)}\beta(1-\beta)c$，$v_2 = v_0$的根为$\delta = 1 - (1-k)(1-r)\beta(1-\beta)c$。

因此，根据边界的连续性，我们得到：只有当 $\max\{\delta_1, 0\} < \delta < \min\{\delta_0, \delta_2\}$ 时，披露政策才能实现更高的利润。其中，δ_0是$v = v_1$的根。当 $0 < \delta < 1 - (1-r)\sqrt{(1-\beta)(k(2-\beta k)-1)}\beta(1-\beta)c$时，$\delta_1$是 $v = v_5$的根。当 $1 - (1-r)\sqrt{(1-\beta)(k(2-\beta k)-1)}\beta(1-\beta)c < \delta < 1 - (1-k)(1-r)\beta(1-\beta)c$时，$\delta_1$是 $v = v_3$的根，δ_2是 $v = v_2$的根。当 $1 - (1-k)(1-r)\beta(1-\beta)c < \delta < 1$时，$\delta_2$是 $v = v_0$的根。 ■

命题5的证明

根据方程9，我们可以推导出两种情况下创作者剩余分别为：

$$CS_N^* = \begin{cases} \frac{c^2(\delta-1)^2+2c(r-1)(-\delta+v(\beta+\delta-1)+1)+(r-1)^2\left(2\beta v+(v-1)^2\right)}{2\beta(1-r)v} & \underline{v_1} < v < \hat{v} \\ \frac{1}{2}(2-\beta)(1-r)v - c\delta & \hat{v} < v < \overline{v}. \end{cases}$$

$$CS_D^* = \begin{cases} \frac{c^2(\delta-1)^2+2c(r-1)(\delta(v-1)+1)+(r-1)^2\left(v^2+1\right)}{2(1-r)v} & \underline{v_1} < v < \min\{\tilde{v}_3, \tilde{v}_4\} \\ \frac{2r\left((\beta k-1)(\beta c(\delta-1)k+1)-v(\beta k-1)(-2\beta+\beta c\delta k+2)+v^2\left(\beta\left(\beta\left(k^2-1\right)-k+2\right)-1\right)\right)}{2\beta k(r-1)v(\beta k-1)} & \\ -\frac{(\beta k-1)(\beta c(\delta-1)k(c(\delta-1)+2)+1)+2v(\beta k-1)(-\beta+\beta c\delta k+1)}{2\beta k(r-1)v(\beta k-1)} & \\ +\frac{r^2\left(v^2\left(\beta\left(\beta-\beta k^2+k-2\right)+1\right)-\beta k-2(\beta-1)v(\beta k-1)+1\right)+v^2\left(\beta\left(\beta-\beta k^2+k-2\right)+1\right)}{2\beta k(r-1)v(\beta k-1)} & \tilde{v_3} < v < \tilde{v}_5 \\ \frac{(1-r)v\left(\beta^2-\beta(k+2)+2\right)}{2-2\beta k} & \tilde{v}_4 < v < \tilde{v}_3 || \max\{\tilde{v}_3, \tilde{v_5}\} < v < \overline{v}. \end{cases}$$

我们比较两种情况下的创作者剩余 $\Delta CS = CS_N^* - CS_D^*$：

(i) When $\underline{v}_1 < v < \underline{v}_2$ and $\underline{v}_2 < v < \min\{\hat{v}, \tilde{v}_3\}$, $\Delta CS = \frac{(1-\beta)(1+\delta c-c+rv-r-v)^2}{2\beta(1-r)v} > 0$

(ii) When $\tilde{v}_3 < v < \hat{v}$, $\Delta CS = \frac{(k-1)(r-1)^2(\beta k-(\beta-1)v(-2\beta k+v(\beta+\beta k-1)+2)-1)}{2\beta k(1-r)v(\beta k-1)} - \frac{2(\beta-1)c(\delta-1)(r-1)(v-1)-(\beta-1)c^2(\delta-1)^2}{2\beta(1-r)v} > 0.$

(iii) 当 $\hat{v} < v < \min\{\tilde{v}_3, \tilde{v}_4\}$, $\Delta CS = \frac{2c(1-r)(1-\delta)+(1-r)^2 1-\beta v^2-1-c^2 1-\delta^2((\,)\,)(\,)(\,)v}{2\ 1-r} > 0$ 时。

(iv) 当 $\max\{\hat{v}, \tilde{v}_3\} < v < \tilde{v}_5$ 时，

$$\Delta CS = \frac{\beta c(\delta-1)k(c(\delta-1)+2)-2r(\beta c(\delta-1)k+1)+r^2+1}{2\beta k(r-1)v} + \frac{2(1-\beta)(1-r)^2}{2\beta k(1-r)} + \frac{(\beta-1)(r-1)^2 v(\beta(k(\beta k-1)-1)+1)}{2\beta k(1-r)(1-\beta k)} > 0$$

(v) 当 $\tilde{v}_4 < v < \tilde{v}_3 || \max\{\tilde{v}_3, \tilde{v}_5\} v < v$ 时，$\Delta CS = \frac{(1-\beta)\beta(1-k)(1-r)_{2-2\beta k}}{v} > 0$。

因此，$CS_N \geq CS_D^*$ 始终成立。 ■

命题6的证明

根据方程10，我们可以推导出两种情况下的创作者剩余分别为：

$$T_N^* = \begin{cases} \frac{(1-r)(1-\beta+(2\beta-1)v)-(1-\beta)(1-\delta)c}{\beta(1-r)v} & \underline{v}_1 < v < \hat{v} \\ \beta & \hat{v} < v < \overline{v}. \end{cases}$$

$$T_D^* = \begin{cases} \frac{1-v-\beta^2(k(kv-1)+v)+\beta((k+2)v-k-1)}{\beta kv(1-\beta k)} & \tilde{v_3} < v < \tilde{v}_5 \\ \frac{\beta(2-\beta-k)}{2-\beta k} & \tilde{v}_4 < v < \tilde{v}_3 || \max\{\tilde{v}_3, \tilde{v_5}\} < v < \overline{v}. \end{cases}$$

我们比较两种情况下平台的内容透明度 $\Delta T = T_N^* - T_D^*$

(i) 当 $v_1 < v < v_2$和 $v_2 < v < \min\{\hat{v}, \tilde{v}_3\}$, $\Delta T = \frac{(1-\beta)((1-r)(1-v)-c(1-\delta))\beta(1-r)}{v} < 0$(ii) 当 $\tilde{v}_3 < v < \hat{v}$,

$$\Delta T = \frac{(\beta-1)(c(\delta-1)+r-1)}{\beta k(r-1)v} + \frac{v\left(\beta\left(-\beta\left(k^2+1\right)+k+2\right)-1\right)}{\beta kv(\beta k-1)} + \frac{-\beta+(2\beta-1)v+1}{\beta v} < 0$$

(iii) 当 $\hat{v} < v < \min\{\tilde{v}_3\, \tilde{v}_4\}$, $\Delta T = \beta - 1 < 0$。

(iv) 当 $\max\{\hat{v}, \tilde{v}_3\} < v < \tilde{v}_5$,

$$\Delta T = \frac{(1-\beta)\left(\beta^2k^2v+\beta k-\beta(k+1)v+v-1\right)}{\beta kv(1-\beta k)} < 0$$

(v) 当 $\tilde{v}_4 < v < \tilde{v}_3 || \max\{\tilde{v}_3, \tilde{v}_5\} < v < v$, $\Delta T = -\frac{(1-\beta)(1-k\beta)}{1-\beta k} < 0$。

因此, $T_N^* \le T_D^*$始终成立。

根据方程 11，我们可以推导出两种情况下的创作者剩余分别为:

$$Q_N^* = \begin{cases} \frac{c(\delta-1)(v-1)+(r-1)(v(\beta+v-2)+1)}{\beta(r-1)v} & \underline{v_1} < v < \hat{v} \\ v & \hat{v} < v < \overline{v}. \end{cases}$$

$$Q_D^* = \begin{cases} \frac{c(\delta-1)(v-1)}{(r-1)v} + v + \frac{1}{v} - 1 & \underline{v}_1 < v < \min\{\tilde{v}_3, \tilde{v}_4\} \\ \frac{\beta ck+\beta c\delta k(v-1)-v(\beta+\beta ck+\beta(k-1)v+v-2)+rv(\beta+\beta(k-1)v+v-2)+r-1}{\beta k(r-1)v} & \tilde{v}_3 < v < \tilde{v}_5 \\ v & \tilde{v}_4 < v < \tilde{v}_3 || \max\{\tilde{v}_3, \tilde{v}_5\} < v < \overline{v}. \end{cases}$$

我们比较两种情况下的内容质量 $\Delta Q = Q_N^* - Q_D^*$

(i) 当 $v_1 < v < \min\{\hat{v}, \tilde{v}_3\}$ 时, $\Delta Q = (\beta-1)(v-1)(c(\delta-1)+(r-1)(v-1))\beta(1-r)v$，这是关于 v 的凸函数，且两个根分别为 $v_6 = \frac{1-r-c}{(1-\delta)(1-r)}$ 和 $v_7 = 1$。$v_6 < v_1$。因此，我们可以推导出：当且仅当 $v_1 < v < \min\{\hat{v}, 1\}$ 时, $\Delta Q < 0$；当且仅当 $1 \le v < \min\{\hat{v}, \tilde{v}_3\}$ 时, $\Delta Q > 0$。

(ii) 当 $\tilde{v}_3 < v < \hat{v}$ 时,

$$\Delta Q = \frac{(v-1)(k((\beta-1)c(\delta-1)+(r-1)((\beta-1)v+1))-(r-1)((\beta-1)v+1))}{\beta k(1-r)v} > 0.$$

(iii) 当 $\hat{v} < v < \min\{\tilde{v}_3, \tilde{v}_4\}$ 时, $\Delta Q = \frac{(v-1)(1-r-c(1-\delta)(1-r)v}{}$。因此, 当且仅当 $1 < v < \min\{\tilde{v}_3, \tilde{v}_4\}$ 时, $\Delta Q > 0$; 当且仅当 $\hat{v} < v < \min\{\tilde{v}_4, 1\}$ 时, $\Delta Q < 0$。

(iv) 当 $\max\{\hat{v}, \tilde{v}_3\} < v < \tilde{v}_5$ 时,

$$\Delta Q = \frac{(v-1)(1-r-v-\beta ck+\beta(c\delta k+v)+(1-\beta)rv)}{\beta k(1-r)v} > 0.$$

(v) 当 $\tilde{v}_4 < v < \tilde{v}_3 || \max\{\tilde{v}_3, \tilde{v}_5\} < v < v$ 时, $\Delta Q = 0$。

因此, 当且仅当 $v_1 < \min\{\tilde{v}_4, 1\}$ 时, $Q_N^* < Q_D^*$; 当且仅当 $1 < v < \tilde{v}_4 || 1 < v < \tilde{v}_5$ 时, $Q_N^* > Q_D^*$。否则, $\Delta Q = 0$ ■