December 2024

AI and Human Creativity on Content Creation: Analyzing Watermark and Detection Policies

Zhenhua Wu
School of Business, Nanjing University, wu.nju@hotmail.com

Lin Hu
Australian National University, lin.hu@anu.edu.au

Pei-yu Chen
ARIZONA STATE UNIVERSITY, peiyu.chen@asu.edu

AI and Human Creativity on Content Creation: Analyzing Watermark and Detection Policies

Completed Research Paper

Zhenhua Wu
School of Business, Nanjing University
Nanjing, China
wu.nju@hotmail.com

Lin Hu
College of Business and Economics, Australian National University
Canberra ACT 2601, Australia
lin.hu@anu.edu.au

Pei-yu Chen
W. P. Carey School of Business, Arizona State University
Tempe, AZ, United States
peiyu.chen@asu.edu

Abstract

This paper investigates the dynamic interaction between artificial intelligence (AI) and human creativity on digital content platforms, focusing on how AI affects content creators' incentives and consumer welfare through content quality. We develop a model to analyze AI's impact, particularly its substitutive and complementary relationships with human skills, and examine the effectiveness of AI watermarking and detection policies. Our findings indicate that AI benefits low-skill creators when human-input costs are low but may push high-skill creators out of the market when costs are high, reducing overall content quality. AI watermarking, which marks AI-generated content, supports high-skill creators if human-input costs are moderate, thus enhancing content quality. Both watermarking and AI detection services improve content quality and consumer satisfaction, with the effectiveness depending on watermark removal rates. High-skill creators prefer watermarking when removal rates are high and detection services when they are low, benefiting consumers and platforms alike.

Keywords:

Artificial Intelligence (AI), AI watermarking & detection, human skill, AI substitutive, AI complementary, content creator

Introduction

The rise of Artificial Intelligence (AI) has sparked a crucial debate within the labor market. Concerns have emerged regarding its potential to displace human content creators, such as AI-powered music generation and automated video editing. These AI-powered tools raise questions about the future of jobs traditionally held by humans, see (Manyika et al., 2017). However, the reality might be more nuanced. AI is currently recognized as a general-purpose technology, as discussed by McAfee and Brynjolfsson (2017). AI may serve as both a complement and a substitute for human skills, see Agrawal et al. (2019). For instance, (De Cremer and Kasparov, 2021) argues that AI complements human skills and is used to augment human intelligence, leading to a more effective and creative workforce. Meanwhile, AI algorithms can analyze vast datasets, identify potential drug candidates, and predict their efficacy. For example, DeepMind's AlphaFold has made significant strides in protein folding prediction, which is crucial for drug discovery. By automating this process, AI substitutes for human labor, accelerating drug development. see (Jumper et al., 2021; Callaway, 2022).

Concurrently, as generative AI models transition from research prototypes to commercial products, they present unprecedented opportunities for content creation, enabling creators to produce vast amounts of content easily. However, with the automation of AI creation processes, there is a risk that creators may overly rely on AI-generated templates or formulas, leading to a homogenization of content and a decrease in originality. Moreover, the ease of content creation facilitated by generative AI may result in the displacement of creators with original ideas by those who lack creativity and solely use AI to generate content. As a result, the quality of content delivered to consumers will depend on the types of creators in the market and how AI is utilized, irrespective of whether AI serves as a substitute or complement to human labor. Therefore, it brings us to our first research question: Given the substitutive and complementary nature of AI and human skills, how does the presence of AI technology affect content creators' entry decisions and usage of AI, as well as consumer welfare from the perspective of content quality?

In response to concerns about content credibility and the ease of producing potentially harmful material, as documented by Simon et al. (2023), policymakers have shifted their focus to distinguishing AI-generated content from human-authored content. Among the proposed solutions, "digital watermarking" has emerged as a promising approach. For instance, in October 2023, President Joe Biden signed an Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, which includes provisions related to AI watermarking, see The White House (2023). While digital watermarking offers a layer of protection, it is not foolproof. Malicious actors can employ techniques such as signal processing or cropping to remove or alter watermarks, making it difficult to definitively identify the origin of content, as discussed by (Singh et al., 2013; Brummer, 2024). Furthermore, each AI model developer can only build detectors for their own watermark, reducing the efficiency of digital watermarking usage. In light of these technological options and their limitations, another question we try to address in this paper is: Do regulatory policies such as AI watermarking and detection effectively protect creators with high human skill levels and improve content quality?

To this end, we develop an analytical model in which a content creator, with or without human skills, can utilize AI to provide content for a consumer on the content platform. The consumer values high-quality content and wishes to support content creators with human skills. We investigate how the introduction of AI technology affects the creator's incentives and consumer welfare from the perspective of content quality. Furthermore, we examine the impact of contentious policies such as AI watermarking and detection on both content creators and consumers.

Our main findings are as follows. Firstly, we find that with the introduction of AI, which can enhance content quality at a lower cost on the supply side, the expected utility of low-skill creators increases when the human input is not too high. However, the expected utility of high-skill creators may decrease, particularly when the human-input cost is higher or when there is a high complementarity between AI and human skills and the human-input cost is low. In extreme cases where the human-input cost is sufficiently high, the presence of AI may lead high-skill creators to exit the market, resulting in a decrease in the average quality of content. On the demand side, if the human-input cost is moderate, AI may reduce the average content quality and the expected quality of subscribed channels. If the human-input cost is low, the presence of AI may decrease consumers' subscription rates, but the average content quality will depend on the balance between the improved content quality due to AI's enhancement of high-skill creators' skills and the reduced content quality due to the entry of low-skill creators. This further implies that AI's contribution to improved content quality increases with the fraction of high-skill creators and AI complementarity, but decreases with human exclusivity.

Secondly, we introduce AI watermarking, which imperfectly removes the footprint of whether AI is used in created content. We find that, on average, the AI watermark policy is beneficial to high-skill creators if the cost of human input is not excessively high, helping prevent their exit from the platform in the presence of AI. However, if AI complementarity is low, the watermark removal rate is high, and the cost of human input is above a certain level, then even with watermarking, high-skill creators might choose to quit the market, indicating the limitations of the watermark policy. Conversely, when AI complementarity is high, the removal rate is low, and the cost of human input is moderate, the AI watermark can effectively distinguish human-created content from AI-generated content, enabling high-skill creators to remain in the market. When the cost of human input is low, the AI watermark boosts the high-skill creator's utility and further enhances it when the watermark removal rate decreases. On the demand side, when AI complementarity is low, the watermark removal rate is high, and the cost of human input is above a certain level, then regardless of the watermark policy, because high-skill creators do not enter the platform, consumers will not subscribe to low-skill creators' channels. Consequently, the policy neither affects the average quality nor the subscription rate. Conversely, when the cost of human input is moderate and high-skill creators stay in the market, the average content quality and the expected quality post-subscription increase. When the cost of human input is low, the AI watermark helps consumers distinguish between different types of creators, thereby enhancing their expected quality for content from subscribed channels.

Thirdly, we consider the scenario where the content platform provides an AI detection service, which determines whether the content was primarily generated by AI or by a human creator. Consumers can then imperfectly observe the outcome of the AI detection due to technological limitations. We find that, similar to the AI watermark, on average, AI detection is beneficial to high-skill creators if the cost of human input is not excessively high, and they remain in the market when AI is introduced. The results from both the supply side and demand side regarding the AI watermark apply here as well.

Finally, when comparing the AI watermark policy to the AI detection service, for high-skill creators, if the AI watermark's removal rate is above a certain level, the AI watermark policy is beneficial to them; if the removal rate is low, the AI detection service is better. On the demand side, the average quality of content is the same across the AI watermark policy and the AI detection service. However, if the removal rate is high, the expected quality will be higher under the AI watermark policy than under the AI detection policy. Otherwise, the AI detection policy results in higher expected quality.

Literature

First of all, our paper has a novel contribution to the expanding research on Artificial Intelligence (AI) and its impact on the labor market. Early studies related to AI, such as Autor et al. (2003), focused on computerization. They found that the automation of routine tasks induced by computation leads to significant changes in the skill demands of the labor market. Brynjolfsson and McAfee (2014) investigate the effects of AI and automation on employment, arguing that while these technologies have the potential to boost productivity and create new job opportunities, they also pose challenges in terms of job displacement and income inequality. Recently, Acemoglu and Restrepo (2018) builds a theoretical model to argue that, in the macro level, the AI can lead to stagnating labor demand, declining labor share in national income, rising inequality, and lower productivity growth (Agrawal et al., 2019) focus on the predictive power of AI, which complements labor in tasks. These studies argue that AI may either increase or decrease the relative return to labor versus capital in decision making. Acemoglu et al. (2022) argue that while AI is substituting for humans in certain tasks, its overall impact on the labor market is not yet detectable.

Our paper contributes to the growing literature on the online "content market". One stream of literature focuses on the supply side of the content market. This research mostly aims to understand how creators can influence user decisions and information diffusion through information technology. For instance, (Li and Du, 2011; Susarla et al., 2012; Claussen et al., 2013; Han et al., 2020) have found that both content creator characteristics and content characteristics affect the diffusion of content. Additionally, Mallipeddi et al. (2021) found that the tone of social media content chosen by the creator and the popularity of the content creator can influence user engagement decisions. On the demand side, research aims to understand how users engage with or share information provided by content creators and further how it affects creators' actions. For instance, Toubia and Stephen (2013) found that consumers' connectivity in their social network can influence the entry of content creators. Berman and Katona (2020) found that users' social connectivity and content quality are strategic complements.

A few recent studies also aim to investigate content production and revenue sharing in the content market, involving all players: content creators, users, and platforms. Tang et al. (2012) studied how different incentives, such as exposure, revenue sharing, and reputation, influence content creators' contributions on social media platforms. Amaldoss et al. (2021) examined the strategies media platforms use, such as consumer fees or advertising, to compensate content creators and how these strategies affect platform profits. Jain and Qian (2021) investigated the impact of creator competition and consumer characteristics on the quality of creators' content and the profits earned by the creators. In Bhargava (2022), it studied how the distribution of creator capabilities affects market concentration among creators and how platform design can influence this distribution. Qian and Jain (2024) focused on the platforms' recommendation strategy and how it might lead to a win-win situation, benefiting the platform, consumers, and content creators.

Our study contributes to the existing literature in the following ways. First, most existing studies focus on the production sector of the economy and attempt to understand how AI affects demand for labor, wages, unemployment, and other macro-level measures. Our work takes a novel approach by studying the impact of AI on the labor market from a micro-level perspective. In our model, we incorporate both substitutes and complements between AI and human skills. Additionally, we allow for heterogeneity and information asymmetry regarding creators' skills, alongside creators competing in the labor market. Our model provides a microfoundation for empirical and macro studies in this field.

Secondly, our research also examines the content market, involving creators, consumers, platforms, and policymakers, which is particularly relevant to the current debate surrounding the use of AI technology and has not been thoroughly studied in the literature. Differing from existing studies, we take an additional step to investigate the market failure caused by the adoption of AI technology and the adverse selection problem on the creator's side. Our results suggest that, under certain conditions, AI adoption may lead to high-skill labor exiting the market, consequently reducing consumer welfare. To mitigate the effects of market failure, we consider the government or platform's implementation of AI watermarking policies or AI detection services. Our findings suggest that appropriately designed policies or services could enhance market efficiency, ensuring the retention of high-skill creators in the market and improving consumer welfare.

The Model

We consider a content platform that includes two key players: a content creator (he) and a content consumer (she). The creator produces content of varying quality, potentially benefiting from the consumer's engagement. The consumer, in turn, rewards the creator for high-quality content by engaging with it.

Content Creator The content creator generates content that can be either high-quality ($q = 1$) or low-quality ($q = 0$). The quality q is influenced by the creator's human skill input $h \in \{0, 1\}$ and the use of AI technology $a \in \{0, 1\}$. The probability of producing high-quality content, $p(h, a) = \Pr(q = 1|h, a)$, is given by: $p(h, a) = \alpha h + \beta \max\{a, h\} + \gamma a h$, where α, β, γ are parameters that capture the distinct effects of human skills, AI technology, and their interaction on content quality. Specifically, α measures human exclusivity, that is, the extent to which the human skill remains beyond the scope of AI substitution and augmentation; β measures AI substitutability, that is, the extent to which the human skills can be substituted by AI; and γ measures AI complementarity, that is, the potential of AI to enhance human skill, improving the quality of content beyond what is possible with human effort alone. These parameters are subject to constraints that ensure the model's realism and relevance: $0 < \alpha, \beta \leq 1$, $0 \leq \gamma \leq 1$, and $\alpha + \beta + \gamma \leq 1$.

To better align our specification with real-world, consider the following examples. Platforms focused on high-quality, professional content, such as Netflix, typically incur higher human-input costs. This is due to significant investments in skills, time, and resources, including hiring experienced writers, directors, and production teams to create content that meets their high standards. In contrast, platforms like TikTok or YouTube often feature user-generated content, where barriers to entry are lower, and human-input costs are relatively minimal. Creators on these platforms can produce and upload videos with basic equipment and minimal editing, leading to lower overall costs.

Similarly, platforms catering to highly specialized content, such as educational platforms like Coursera, often have higher human-input costs. Creators on these platforms must invest in deep expertise and produce well-researched, credible content, which increases the cost of human input. Conversely, platforms that prioritize quantity over quality, or those catering to audiences seeking quick, entertaining content like TikTok, tend to have lower human-input costs. The expectations for content quality are lower, allowing creators to produce content more rapidly and with less investment in professional-grade tools or expertise.

For AI substitutability and complementarity, consider the following examples: Platforms that focus on content requiring a high degree of creativity, originality, or human expertise—such as educational platforms like Coursera—usually experience lower AI complementarity. In these contexts, AI tools are often seen as supplementary, assisting with tasks such as content organization but not replacing the core human-driven content creation. In contrast, platforms that emphasize more formulaic or repetitive content—such as social media platforms like YouTube, TikTok, and Instagram—tend to experience higher AI complementarity. On these platforms, AI can effectively support content creation, recommendations, and even automated editing or enhancement.

When the creator's skill is high ($h = 1$) and AI is not used ($a = 0$), the creator relies solely on traditional methods, utilizing only human effort to produce content. In this scenario, the probability of producing high-quality content is $\alpha + \beta$. Conversely, when the creator lacks human skill ($h = 0$) and depends entirely on AI ($a = 1$), the likelihood of generating high-quality content is β. In situations where both high human skill and AI are employed ($h = a = 1$), AI serves to augment human capability, thereby increasing the likelihood of high-quality content production to $\alpha + \beta + \gamma$. The cost associated with the inputs is $C(a, h) = ch$, implying that AI input is free of charge, but human input incurs a cost $c > 0$.

There are two types of content creators: (i) *high-skill creator*—a creator capable of choosing human skill input $h \in \{0, 1\}$ for content production; this type represents creators who can either utilize or forego human skill inputs in their content creation process; (ii) *low-skill creator*—a creator with inherently low skill level who can only choose human skill level $h = 0$ for content production, thereby relying solely on alternative methods or technologies for content creation. We write $t \in \{H, A\}$ for the creator's type, where H represents the high-skill creator and A represents the low-skill creator. The probability of encountering a high-skill creator is $\Pr(t = H) = \lambda \in (0, 1)$, and for an low-skill creator, it's $\Pr(t = A) = 1 - \lambda$.

The idea behind above specification is that low-skill creators do not invest in human effort, reflecting a fundamental difference in their abilities and production capabilities. In practice, low-skill creators are those who lack the ability or resources to enhance their content through human effort, relying more heavily on tools like AI to remain competitive.

To better illustrate our specification of content creators, we provide additional justification here. First, regarding the concept of "human input," we define it as encompassing all actions undertaken by the content creator that involve substantial human effort. When $h = 1$, this represents innovative actions that are costly due to the learning or accumulation of human capital. This includes activities such as writing prompts and guiding AI. Second, in our definitions of "high-skill" and "low-skill" creators, "high-skill" creators are those who invest considerable human effort, such as acquiring and applying human capital, while "low-skill" creators rely more on AI tools. We do not consider "middle-skill" creators in our model, as the distinction between these two types is sufficient to drive our main findings.

Content Consumer The content consumer prefers high-quality content and takes two kinds of reward actions toward the content creator: (i) short-term actions—engaging with the content via likes, comments, or shares for high-quality finds ($q = 1$); and (ii) long-term actions—subscribing to the creator's channel if she finds the content high quality, with a likelihood $p_s = \Pr(t = H|q = 1)$, indicative of high-quality content perception. Essentially, the probability of subscribing (p_s) reflects the consumer's posterior of the subscriber being a high-skill creator, given the high-skill creator's relatively higher propensity to produce quality content compared to the low-skill creator.

Here, we provide further justification for our consumer specification. First, in our model, the subscription rate is influenced by whether the creator consistently produces high-quality content. This quality is affected by both human input and AI usage. Consequently, the subscription rate can vary continuously within the interval $[0, 1]$. Second, because consumers cannot perfectly observe whether content is generated by AI, their subscription decisions depend solely on the quality of the content. In other words, consumers prefer high-quality content regardless of whether it is produced primarily by AI or human effort. We do not assume that consumers will completely avoid subscribing to content generated with AI.

Payoffs The payoff for a creator from a consumer's positive engagement is normalized to 1, and the payoff from subscriptions is set at $x > 0$.[1] The outside option for any type of content creator is 0. Thus, a creator creates content on the platform if and only if the expected utility of doing so is greater than 0.

For consumers, the payoff from engaging with content is quantified by two metrics: the average current content quality ($E[q]$) and the expected quality of future content contingent upon subscription ($E[q'|s]$), with q' indicating the quality of upcoming content.

Skill Gap and AI's Role A critical aspect of the model is its capacity to investigate whether AI technology narrows or widens the skill gap among content creators. The skill gap is conceptualized as the difference in the probability of producing high-quality content between creators with high and low human skills. In the absence of AI ($a = 0$), the skill gap is $\alpha + \beta$, reflecting the combined influence of innate human skill and the potential for skill substitution by AI. The introduction of AI ($a = 1$) alters this dynamic, making the revised skill gap $\alpha + \gamma$. The change induced by AI, therefore, is $\gamma - \beta$.

This approach allows for a differentiated analysis of AI's impact: When $\beta > \gamma$, AI primarily acts as a substitute for human skill, potentially narrowing the skill gap by providing tools that level the playing field between high- and low-skilled creators. Conversely, if $\beta < \gamma$, AI enhances human skill more significantly than it substitutes it, potentially widening the skill gap by augmenting the capabilities of already skilled creators to a greater extent than it aids less skilled ones.

Analysis

Benchmark: Traditional Technology (without AI)

To evaluate the influence of AI on content creation effectively, we commence with a scenario where AI technology is not employed by creators. In this setting, content is exclusively produced by high-skill creators.

If a high-skill creator chooses human input $h = 1$, his expected utility is $U_H = (\alpha+\beta)(1+x)-c$. If he chooses human input $h = 0$, his expected utility is 0. Thus, he would choose $h = 1$ if and only if $(\alpha + \beta)(1 + x) > c$. To simplify the notations, we let $\bar{c} = (\alpha + \beta)(1 + x)$. If the cost of human input, c, exceeds $\bar{c}$, the expected utility for a high-skill creator choosing $h = 1$ becomes negative, leading to no human input $h = 0$ in content creation. To avoid such trivial cases, we assume $c < \bar{c}$.

Consequently, the probability that a high-skill creator produces high-quality content, denoted as p, equals $\alpha + \beta$ (i.e., $p = \alpha + \beta$).

In this context, consumers presume all content to originate from high-skill creators. The consumer updates her belief and believes that the content creator is high-skill with probability 1, leading the consumer to subscribe to the creator's channel with certainty ($p_s = 1$). Hence, the expected utility for a high-skill creator, operating without AI is $U_H = (\alpha + \beta)(1 + x) - c$. On the other hand, low-skill creators, lacking the capability to produce content in the absence of AI technology, face an expected utility of $U_L = 0$.

Lemma 1 (Traditional Technology Benchmark)**.** *In the absence of AI technology, the following statements hold. 1) The high-skill creator chooses human input $h = 1$ to create content and has expected utility $U_H = (\alpha + \beta)(1 + x) - c$. The low-skill creator does not create content and has the expected utility $U_L = 0$. 2) The average content quality is $E[q] = \alpha + \beta$. The consumer subscribes to the creator's channel with probability $p_s = 1$ and the expected quality of the subscribed channel is $E[q'|s] = \alpha + \beta$.*

The Role of AI in Content Creation

Exploring the incorporation of AI technology into content creation reveals its significant benefits. AI not only aids high-skill creators by amplifying their human skills but also empowers low-skill creators to produce content at no additional cost. Given AI's capability to elevate content quality cost-effectively, it is rational for both high-skill and low-skill creators to adopt AI technology in their content production endeavors. Thus,

[1]For detailed insights into monetization and the impact of engagement, refer to YouTube's Partner Program (https://support.google.com/youtube/answer/72851?hl=en) and further resources on sponsorship and merchandise sales (https://thepennymatters.com/how-much-do-sponsors-pay-youtubers/).

$a = 1$ for $t \in \{H, L\}$.

In the context of AI utilization, the low-skill creator chooses $(h, a) = (0, 1)$, and the associated probability of making high-quality content is $p_L^{AI}(0,1) = \beta$. For the high-skill creator, if he chooses $h = 0$, $p_H^{AI}(0,1) = \beta$; and if he chooses $h = 1$, $p_H^{AI}(1,1) = \alpha + \beta + \gamma$.

Suppose the high-skill creator chooses $h = 0$. The probability of the consumer encountering high-quality content is $p^{AI} = \beta$. Observing high-quality content, the consumer expects the probability that the creator is of high-skill creator is λ, and subscribes to the creator with probability λ. In this case, the expected utility of the high-skill creator is $U_H^{AI}(0,1) = \beta(1 + \lambda x)$. The expected utility of the low-skill creator is $U_L^{AI}(0,1) = \beta(1 + \lambda x)$.

Suppose the high-skill creator chooses $h = 1$. The probability of the consumer encountering high-quality content is

$$p^{AI} = \lambda(\alpha + \beta + \gamma) + (1 - \lambda)\beta = \beta + \lambda(\alpha + \gamma).$$

Observing high-quality content, the consumer expects the probability that the creator is of high skill (subscription probability) is

$$p_s = \Pr(t = h | q = 1) = \frac{\lambda(\alpha + \beta + \gamma)}{\beta + \lambda(\alpha + \gamma)}.$$

In this case, the expected utility of the high-skill creator is

$$U_H^{AI}(1,1) = (\alpha + \beta + \gamma)\left[1 + \frac{\lambda(\alpha + \beta + \gamma)}{\beta + \lambda(\alpha + \gamma)} x\right] - c.$$

The expected utility of the low-skill creator is

$$U_L^{AI}(0,1) = \beta\left[1 + \frac{\lambda(\alpha + \beta + \gamma)}{\beta + \lambda(\alpha + \gamma)} x\right].$$

Therefore, the equilibrium with $h = 1$ in the presence of AI is sustainable if and only if

$$U_H^{AI}(1,1) - U_H^{AI}(0,1) = \alpha + \gamma - \lambda x\left[1 - \frac{(\alpha + \beta + \gamma)^2}{\beta + \lambda(\alpha + \gamma)}\right] - c > 0.$$

To simplify the notations, we let $\bar{c}^{AI} = \alpha + \gamma - \lambda x\left[1 - \frac{(\alpha+\beta+\gamma)^2}{\beta+\lambda(\alpha+\gamma)}\right]$ for the threshold for the high-skill creator to use human input. Then the high-skill creator chooses $h = 1$ if and only if $c < \bar{c}^{AI}$.

Lemma 2 (The choice of human input)**.** *The high-skill creator's choice of human input is determined by the cost comparison with the threshold $\bar{c}^{AI}$: (i) The high-skill creator opts to utilize human input, i.e., $h_H = 1$ if the cost of his human input $c < \bar{c}^{AI}$; (ii) otherwise, the high-skill creator refrains from using human input, i.e., $h_H = 0$. In particular, $\bar{c}^{AI}$ is increasing in AI exclusivity α and AI complementarity γ.*

Lemma 2 characterizes the condition that the high-skill creator still utilizes human input in the presence of AI, that is, when the cost of human input is not very high, and/or AI exclusivity/complementarity is high. In this case, the benefit from using human input exceeds the cost.

We further examine the impact of AI on the creators' utility. First, we consider the case $c \geq \bar{c}^{AI}$. For the high-skill creator, the presence of AI changes his expected utility by

$$\begin{aligned}\Delta U_H^{AI}(h_H = 0) &= U_H^{AI}(0,1) - U_H = \beta(1 + \lambda x) - (\alpha + \beta)(1 + x) + c \\ &= -\alpha(1 + x) - \beta(1 - \lambda)x + c.\end{aligned}$$

For the low-skill creator, the presence of AI changes his expected utility by

$$\Delta U_L^{AI}(h_H = 0) = U_L^{AI}(0,1) - 0 = \beta(1 + \lambda x).$$

Then, we turn to the case $c < \bar{c}^{AI}$. For the high-skill creator, the presence of AI changes his expected utility by

$$\begin{aligned}\Delta U_H^{AI}(h_H = 1) &= U_H^{AI}(1,1) - U_H = \bar{c}^{AI} - \bar{c} \\ &= \underbrace{\gamma}_{\Delta U \text{ from short-term reward}} - \underbrace{\left[\alpha + \beta - \frac{\lambda(\alpha + \beta + \gamma)^2}{\beta + \lambda(\alpha + \gamma)}\right] x}_{\Delta U \text{ from long-term reward}}.\end{aligned}$$

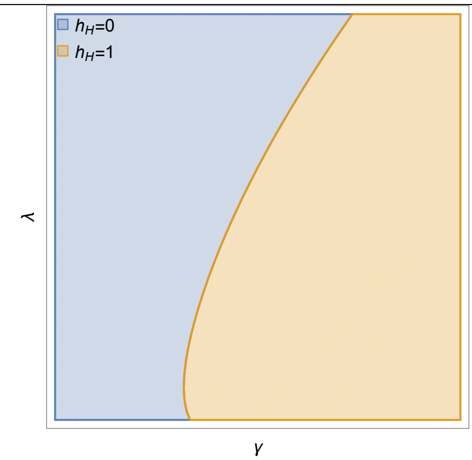

Figure 1. AI and the high-skill creator: $\alpha = 0.1$, $\beta = 0.3$, $x = 1.5$, $c = 0.3$

For the low-skill creator, the presence of AI changes his expected utility by

$$\Delta U_L^{AI}(h_H = 1) = U_L^{AI}(0,1) - 0 = \beta \left[1 + \frac{\lambda(\alpha+\beta+\gamma)}{\beta+\lambda(\alpha+\gamma)}x\right] > \Delta U_L^{AI}(h_H = 0).$$

Proposition 1 explores the economic impact of AI on both high-skill and low-skill creators.

Proposition 1 (AI and Creator)**.** *The effect of AI on the creator is as follows.*

(i) For the low-skill creator, the presence of AI improves his expected utility: the expected gain is higher when $c < \bar{c}^{AI}$*.*

(ii) For the high-skill creator, the presence of AI improves his expected utility if $c \geq \max\{\bar{c}^{AI}, \alpha(1+x) + \beta(1-\lambda)x\}$*, or* $c < \bar{c}^{AI}$ *and* $\gamma > \left[\alpha + \beta - \frac{\lambda(\alpha+\beta+\gamma)^2}{\beta+\lambda(\alpha+\gamma)}\right]x$*; otherwise, the presence of AI decreases his expected utility.*

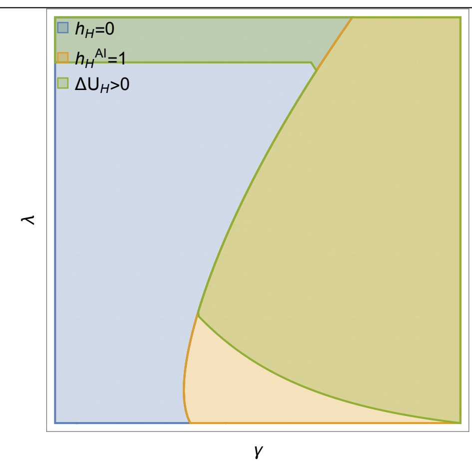

Figure 2. AI and the high-skill creator:$\alpha = 0.1$, $\beta = 0.3$, $x = 1.5$, $c = 0.3$

Next, we turn to how the introduction of AI affects the consumer in terms of content quality. If $c \geq \bar{c}^{AI}$, the high-skill creator does not use human input $h_H = 0$ and the average quality is $E[q] = \beta$; the expected quality for the subscribed channel with AI is $E[{q'}^{AI}|s] = \beta$. If $c < \bar{c}^{AI}$, the high-skill creator utilizes human input $h_H = 1$ and the average quality is $E[q] = \beta + \lambda(\alpha + \gamma)$; the expected quality for the subscribed channel with

AI is

$$E[{q'}^{AI}|s] = \mathbf{Pr}(t = H|q = 1)(\alpha + \beta + \gamma) + \mathbf{Pr}(t = A|q = 1)\beta$$
$$= \beta + \frac{\lambda(\alpha + \gamma)(\alpha + \beta + \gamma)}{\beta + \lambda(\alpha + \gamma)}.$$

We then evaluate how the presence of AI influences the consumer from the perspective of content quality. Proposition 2 presents the results.

Proposition 2 (AI and consumer)**.** *The effect of AI on the consumer is as follows.*

(i) If $\bar{c}^{AI} < c < \bar{c}$, the presence of AI reduces the consumer's subscription probability from 1 to λ. It reduces the average content quality and the expected quality of subscribed channel by α.

(ii) If $c < \bar{c}^{AI}$, the presence of AI reduces the consumer's subscription probability from 1 to $\frac{\lambda(\alpha+\beta+\gamma)}{\beta+\lambda(\alpha+\gamma)}$. It changes the average quality of the content for the consumer by $\Delta E[q^{AI}] = \lambda\gamma - (1-\lambda)\alpha$, and changes the expected quality for the subscribed channel by $\Delta E[{q'}^{AI}|s] = \frac{\lambda(\alpha+\gamma)(\alpha+\beta+\gamma)}{\beta+\lambda(\alpha+\gamma)} - \alpha$.

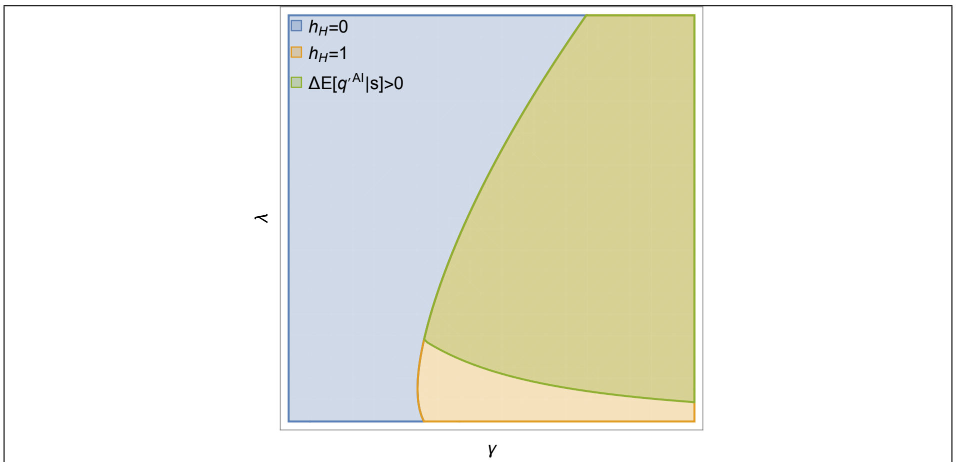

Figure 3. AI and the high-skill creator: $\alpha = 0.1$**,** $\beta = 0.3$**,** $x = 1.5$**,** $c = 0.3$

By Proposition 1, when the cost of human skill input is sufficiently high, the introduction of AI induces the high-skill creator to exit from the platform. As a consequence, the average quality of the content on the platform decreases, and the consumer would not subscribe to the creator's channel.

Conversely, if the cost of human skill input is not that high and the high-skill creator stays in the platform, the introduction of AI increases (respectively, decreases) the average content quality when the improved content quality by AI's enhancement on the high-skill creator's skill, $\lambda\gamma$, is greater (respectively, smaller) than the reduced content quality by the entry of the low-skill creator, $(1 - \lambda)\alpha$. Therefore, AI's contribution to the improved content quality is increasing in the fraction of high-skill creators, λ, and AI complimentarity, γ; it is decreasing in the human exclusivity, α.

Notably, the expected quality of the subscribed channel is increasing in the likelihood that the high-quality content is generated by a high-skill creator, which is increasing in the fraction of high-skill creators λ, AI complimentarity γ, and decreasing in AI substitutability β. As a result, AI's contribution to the improved expected quality of the subscribed channel is increasing in the fraction of high-skill creators λ, AI complimentarity γ, and decreasing in AI substitutability β. Moreover, the presence of AI induces the low-skill creator who does not have human skill and is not affected by human exclusivity α. Thus, α plays a less important role in the presence of AI. Therefore, AI's contribution to the improved expected quality of the subscribed channel is decreasing in the human exclusivity α.

AI Watermark and Detection Policies

While AI augments human productivity, it concurrently introduces low-skill creators that could undermine the value of human contributions. This dynamic poses challenges for tasks necessitating genuine human input, sparking debates on the implementation of policies like watermarking or AI detection to safeguard human contributions. This section studies the implications of such policies on the incentives of creators and the welfare of consumers.

AI Watermark Policy

We explore a scenario in which AI-generated content is identifiable by a watermark, signifying its AI origin. Consumers can discern whether content has been primarily produced by AI based on the presence of this watermark. However, due to technological constraints, there exists a possibility that the AI watermark might be successfully removed, allowing some AI-generated content to appear unmarked. To quantify this, we introduce $\rho \in (0,1)$ as the watermark removal rate—the probability that a piece of AI-generated content will lack the watermark. We also define $m \in \{0,1\}$ to represent the AI watermark status, with $m = 1$ denoting content identified as AI-generated, and $m = 0$ denoting content without such identification.

Content that involves human input, not being primarily generated by AI, will not carry the AI-generated watermark. Therefore, any content produced by high-skill creators will inherently be free from being marked as AI-generated.

With the implementation of an AI watermark policy, conditional on both types of creators entering the platform, the overall probability of encountering high-quality content remains unaffected, described as:

$$p^W = \beta + \lambda(\alpha + \gamma).$$

However, the probability that a consumer subscribes to a channel has been altered due to the watermark's influence. For content not marked as AI-generated($m = 0$), the consumer's posterior belief, given high-quality content ($q = 1$) is

$$p_s^W(m = 1) = \Pr(t = H|q = 1, m = 1) = 0,$$

indicating that if the content is marked as AI-generated, the consumer will not subscribe, believing it lacks human contribution. Conversely, for content not marked as AI-generated, the consumer's posterior belief, given high-quality content ($q = 1$) is

$$p_s^W(m = 0) = \Pr(t = H|q = 1, m = 0) = \frac{\lambda(\alpha + \beta + \gamma)}{\lambda(\alpha + \beta + \gamma) + (1 - \lambda)\beta\rho},$$

highlighting a higher likelihood of subscription under the assumption that unmarked content has a greater chance of being high-skill.

The high-skill creator's expected utility with the AI watermark policy is

$$U_H^W = p(H)(1 + p_s(m = 0)x) = (\alpha + \beta + \gamma)\left[1 + \frac{\lambda(\alpha + \beta + \gamma)x}{\lambda(\alpha + \beta + \gamma) + (1 - \lambda)\beta\rho}\right] - c.$$

The low-skill creator's expected utility with the AI watermark policy is

$$U_A^W = p(A)(1 + \rho p_s(m = 0)x) = \beta\left[1 + \frac{\rho\lambda(\alpha + \beta + \gamma)x}{\lambda(\alpha + \beta + \gamma) + (1 - \lambda)\beta\rho}\right].$$

Proposition 3 details the impact of implementing an AI watermark policy on the expected utilities of various creator types. For simplicity in notation, we define the cutoff cost, $\bar{c}^W$, as follows:

$$\bar{c}^W = (\alpha + \beta + \gamma)\left[1 + \frac{\lambda(\alpha + \beta + \gamma)x}{\lambda(\alpha + \beta + \gamma) + (1 - \lambda)\beta\rho}\right].$$

Should the cost of human skill c exceed $\bar{c}^W$, the expected utility for the high-skill creator, U_H^W, becomes negative, prompting his exit from the platform.

Proposition 3 (AI Watermark and Creator)**.** *The impact of AI watermark policy on the creator is described as follows.*

(i) For the low-skill creator, AI watermark policy decreases his expected utility by

$$\Delta U_A^{AI} = \frac{\lambda(1 - \rho)(\alpha + \beta + \gamma)}{(\beta(1 - (1 - \lambda)(1 - \rho)) + \lambda(\alpha + \gamma))(\beta + \lambda(\alpha + \gamma))}$$

and ΔU_A^{AI} is increasing in the watermark removal rate ρ.

(ii) For the high-skill creator, $\bar{c}^{AI} < \bar{c}^W$. That is, if the creator enters the platform without AI watermark policy, the creator must enter the platform with AI watermark policy. Specifically, the following statements hold.

(a) If $\bar{c}^W < c < \bar{c}$, the high-skill creator exits the content platform regardless of the AI watermark policy. This case exists if and only if $\gamma < (\alpha+\beta)x$ and $\rho > \frac{\lambda\gamma(\alpha+\beta+\gamma)(1+x)}{\beta(1-\lambda)((\alpha+\beta)x-\gamma)}$.

(b) If $\bar{c}^{AI} < c < \bar{c}^W$, AI watermark policy prevents the high-skill creator from exiting the platform; it improves the high-skill creator's expected utility by $\Delta U_H^W = \bar{c}^W - c$, and ΔU_H^W is decreasing in ρ.

(c) If $c < \bar{c}^{AI}$, the high-skill creator enters the platform with or without the policy. AI watermark policy increases his expected utility by $\Delta U_H^W = \lambda x(\alpha+\beta+\gamma)^2\left[\frac{1}{\beta(1-(1-\lambda)(1-\rho))+(\alpha+\lambda)\gamma} - \frac{1}{\beta+(\alpha+\lambda)\gamma}\right]$, and ΔU_H^W is decreasing in ρ.

The AI watermark policy aids consumers in distinguishing high-skill creators from low-skill ones. As the technology advances—specifically as the watermark removal rate ρ decreases—the effectiveness of this differentiation improves, leading to a decrease in the subscription probability for the low-skill creator. In the ideal scenario where watermark technology is perfect, all AI-generated content is marked, and consumers will avoid subscribing to the low-skill creator's channel.

Generally, the AI watermark policy is beneficial to the high-skill creator if the cost of human skill is not excessively high, helping prevent his exit from the platform in the presence of AI. However, there are circumstances where the policy may not offer adequate protection. Specifically, if AI complementarity γ is low and the watermark removal rate ρ is high, the cost of human skill c could exceed the threshold $\bar{c}^W$. In such cases, even with watermarking, the high-skill creator might still find it unfeasible to enter the platform, indicating the limitations of the watermark policy under certain conditions.

Conversely, as γ increases and ρ decreases, a more nuanced situation emerges when the cost of human skill is moderate ($\bar{c}^{AI} < c < \bar{c}^W$). Here, a high-skill creator, who otherwise avoids the platform due to the presence of the low-skill creator, is encouraged to participate thanks to the AI watermarking policy. This policy effectively distinguishes human-created content from AI-generated content, enabling the high-skill creator to maintain a presence on the platform.

When the cost of human skill is low ($c < \bar{c}^{AI}$), the high-skill creator remains competitive in the presence of AI. In this scenario, the AI watermark policy not only boosts the high-skill creator's utility but also enhances it further as the watermark removal rate ρ decreases. This improvement is because consumers can more reliably identify and prefer content produced by the high-skill creator, increasing his utility on the platform.

Suppose the high-skill creator enters the platform. The average quality is

$$E[q^W] = \lambda(\alpha+\beta+\gamma) + (1-\lambda)\beta = \beta + \lambda(\alpha+\gamma)$$

Observing no watermarking, the consumer subscribes to the channel with probability $p_s^W(m=0)$. If subscribed, the expected quality is

$$\begin{aligned} E[q'^W|s, m=0] &= \Pr(t=H|q=1, m=0)(\alpha+\beta+\gamma) + \Pr(t=A|q=1, m=0)\beta \\ &= \beta + \frac{\lambda(\alpha+\gamma)(\alpha+\beta+\gamma)}{\lambda(\alpha+\beta+\gamma)+(1-\lambda)\beta\rho}. \end{aligned}$$

We then can evaluate how AI watermark policy influences the consumer from the perspective of content quality. The results are listed in Proposition 4.

Proposition 4 (AI Watermark and Consumer)**.** *The impact of AI watermarking policy on the consumer is as follows.*

(i) If $\bar{c}^W < c < \bar{c}$, AI watermark policy does not affect the content quality, and the consumer does not subscribe to the creator's channel independent of the policy.

(ii) If $\bar{c}^{AI} < c < \bar{c}^W$, AI watermark policy increases the average quality by $\lambda(\alpha+\gamma)$ and increases the subscription probability from 0 to $\frac{\lambda(\alpha+\beta+\gamma)}{\lambda(\alpha+\beta+\gamma)+(1-\lambda)\beta\rho}$ with expected quality $\beta + \frac{\lambda(\alpha+\gamma)(\alpha+\beta+\gamma)}{\lambda(\alpha+\beta+\gamma)+(1-\lambda)\beta\rho}$.

(iii) If $c < \bar{c}^{AI}$, AI watermark policy does not affect the average content quality, but increases the expected

quality for the subscribed channel by

$$\Delta E[q'^{W}|s] = \left[\frac{1}{\beta(1-(1-\lambda)(1-\rho))+\lambda(\alpha+\gamma)} - \frac{1}{\beta+(\alpha+\gamma)\lambda}\right]\lambda(\alpha+\gamma)(\alpha+\beta+\gamma).$$

The improved expected quality $\Delta \mathrm{E}[q'^{W}|s]$ *is decreasing in the watermark removal rate* ρ.

As indicated in Proposition 4, under conditions where AI complementarity γ is low and the watermark removal rate ρ is high, the cost of human skill c might exceed the threshold $\bar{c}^W$. In such cases, regardless of the watermark policy, the high-skill creator does not enter the platform, and then the consumer does not subscribe to the low-skill creator's channel. Consequently, the policy neither affects the average quality nor the subscription. Conversely, when the cost of human skill is moderate ($\bar{c}^{AI} < c < \bar{c}^W$), the high-skill creator is incentivized to participate thanks to the AI watermark policy. This policy significantly enhances the average content quality and the expected quality post-subscription, as it encourages a richer mix of high-skill content. In situations where the cost of human skill is low ($c < \bar{c}^{AI}$), high-skill creators remain on the platform irrespective of the watermark policy. Here, the watermark policy does not alter the mix of creators and hence does not impact the average content quality of the platform. However, it plays a crucial role in helping consumers more effectively distinguish between different types of creators, thereby enhancing their expected quality for content from subscribed channels.

AI Detector

We now consider the case where the content platform provides the AI detection service. This service detects whether the content was primarily generated by AI or by human. The consumer can then observe the outcome of the AI detection. However, due to technological limitations, we assume that there is a chance of error in the detection method, implying that some AI-generated/human-generated content might not be flagged as such. To quantify this, we introduce $\phi \in [0, \frac{1}{2}]$ as the detection error rate—the probability that any given piece of AI-generated (human-generated) content is detected as human-generated (AI-generated).[2] We write $d \in \{H, A\}$ for the detection outcome, where $d = H$ represents the content detected as human-generated and $d = A$ represents the content detected as AI-generated.

With the implementation of an AI detector, conditional on both types of creators entering the platform, the overall probability of encountering high-quality content remains unaffected, described as: $p^D = \beta + \lambda(\alpha+\gamma)$.

However, the probability that a consumer subscribes to a channel has been altered due to the influence of the AI detector. Now for content not detected as human-generated($d = H$), the consumer's posterior belief, given high-quality content ($q = 1$) is

$$p_s^D(d=H) = \Pr(t=H|q=1, d=H) = \frac{\lambda(\alpha+\beta+\gamma)(1-\phi)}{\lambda(\alpha+\beta+\gamma)(1-\phi)+(1-\lambda)\beta\phi},$$

which is decreasing in the detection error rate ϕ. For content not detected as AI-generated($d = H$), the consumer's posterior belief, given high-quality content ($q = 1$) is

$$p_s^D(d=A) = \Pr(t=H|q=1, d=A) = \frac{\lambda(\alpha+\beta+\gamma)\phi}{\lambda(\alpha+\beta+\gamma)\phi+\beta(1-\lambda)(1-\phi)},$$

which is increasing in the detection error rate ϕ.

The high-skill creator's expected utility with AI detector is

$$\begin{aligned} U_H^D &= (\alpha+\beta+\gamma)[1+(1-\phi)p_s(d=H)x+\phi p_s(d=A)x]-c \\ &= (\alpha+\beta+\gamma) + \frac{\lambda(\alpha+\beta+\gamma)^2(1-\phi)^2 x}{\lambda(\alpha+\beta+\gamma)(1-\phi)+(1-\lambda)\beta\phi} + \frac{\lambda(\alpha+\beta+\gamma)^2\phi^2 x}{\lambda(\alpha+\beta+\gamma)\phi+(1-\lambda)\beta(1-\phi)} - c. \end{aligned}$$

The low-skill creator's expected utility with AI detector is

$$\begin{aligned} U_A^D &= \beta[1+\phi p_s(d=H)x+(1-\phi)p_s(d=A)x] \\ &= \beta + \frac{\beta\lambda(\alpha+\beta+\gamma)(1-\phi)\phi x}{\lambda(\alpha+\beta+\gamma)(1-\phi)+(1-\lambda)\beta\phi} + \frac{\beta\lambda(\alpha+\beta+\gamma)\phi(1-\phi)x}{\lambda(\alpha+\beta+\gamma)\phi+\beta(1-\lambda)(1-\phi)} \end{aligned}$$

Proposition 5 summarizes how AI detector policy affects the expected utilities of different types of creators.

[2] If $\phi > \frac{1}{2}$, AI-generated (human-generated) content will be more likely detected as human-generated (AI-generated) than AI-generated (human-generated), and such technology should never be adopted.

For simplicity in notation, we define the cutoff cost, $\bar{c}^D$, as follows:

$$\bar{c}^D = (\alpha+\beta+\gamma) + \frac{\lambda(\alpha+\beta+\gamma)^2(1-\phi)^2 x}{\lambda(\alpha+\beta+\gamma)(1-\phi)+(1-\lambda)\beta\phi} + \frac{\lambda(\alpha+\beta+\gamma)^2\phi^2 x}{\lambda(\alpha+\beta+\gamma)\phi+(1-\lambda)\beta(1-\phi)}.$$

Should the cost of human skill c exceed $\bar{c}^D$, the expected utility for the high-skill creator, U_H^D, becomes negative, prompting his exit from the platform.

Proposition 5 (AI detector and the creator)**.** *The impact of AI detector on the creator is described as follows:*

(i) For the low-skill creator, AI detector decreases his expected utility. As the detection error rate ϕ increases, the loss in low-skill creator's expected utility decreases.
(ii) For the high-skill creator, $\bar{c}^{AI} < \bar{c}^D$. That is, if the creator enters the platform without AI detection policy, the creator must enter the platform with AI detection policy. Specifically, we have the following:
  (a) If $\bar{c}^D < c < \bar{c}$, the high-skill creator exits the content platform with or without the AI watermark policy. This case exists if γ is sufficiently small and ϕ is sufficiently large.
  (b) If $\bar{c}^{AI} < c < \bar{c}^D$, AI watermark policy prevents the high-skill creator from exiting the platform; it improves the high-skill creator's expected utility by $\Delta U_H^D = \bar{c}^D - c$, and ΔU_H^D is decreasing in ϕ.
  (c) If $c < \bar{c}^{AI}$, the high-skill creator enters the platform with or without the policy. AI watermark policy increases his expected utility by $\Delta U_H^D = \bar{c}^D - \bar{c}^{AI}$, and ΔU_H^D is decreasing in ϕ.

Similar to the AI watermark policy, the AI detector policy also helps consumers distinguish the high-skill creator from the low-skill. As the technology advances—specifically as the detection error rate ρ decreases—the effectiveness of this differentiation improves, leading to a decrease in the subscription probability for the low-skill creator. In the ideal scenario where detection technology is perfect, the content is detected as AI-/human-generated correctly, and consumers will avoid subscribing to the low-skill creator's channel.

Generally, the AI detection policy is beneficial to the high-skill creator if the cost of human skill is not excessively high, helping prevent his exit from the platform in the presence of AI.

Similar to the AI watermark policy, when AI complementarity γ is low and the detection error rate ϕ is high, the cost of human skill c could exceed the threshold $\bar{c}^W$. In such cases, even with AI detector, the high-skill creator might still find it unfeasible to enter the platform, indicating the limitations of the AI detector policy under certain conditions.

Conversely, as γ increases and ϕ decreases, a more nuanced situation emerges when the cost of human skill is moderate ($\bar{c}^{AI} < c < \bar{c}^D$). Here, a high-skill creator, who otherwise avoids the platform due to the presence of the low-skill creator, is encouraged to participate thanks to the AI detector policy. This policy effectively distinguishes human-created content from AI-generated content, enabling the high-skill creator to maintain a presence on the platform.

When the cost of human skill is low ($c < \bar{c}^{AI}$), the high-skill creator remains competitive in the presence of AI. In this scenario, the AI detector policy not only boosts the high-skill creator's utility but also enhances it further as the detection error rate ϕ decreases. This improvement is because consumers can more reliably identify and prefer content produced by the high-skill creator, increasing his utility on the platform.

Suppose any type of the creator enters the platform. The average quality is $\mathrm{E}[q^D] = \beta + \lambda(\alpha+\gamma)$. If the content is detected as human-generated, the consumer subscribes to the channel with probability $p_s^D(d = H)$. If subscribed in this case, the expected quality is

$$\begin{aligned}\mathrm{E}[q^{D'}|s, d=H] &= \mathbf{Pr}(t=H|q=1, d=H)(\alpha+\beta+\gamma) + \mathbf{Pr}(t=A|q=1, d=H)\beta \\ &= \beta + \frac{\lambda(\alpha+\gamma)(\alpha+\beta+\gamma)(1-\phi)}{\lambda(\alpha+\beta+\gamma)(1-\phi)+(1-\lambda)\beta\phi}.\end{aligned}$$

Similarly if the content detected as AI-generated, the consumer subscribes to the channel with probability

$p_s^D(d=H)$. If subscribed in this case, the expected quality is

$$\begin{aligned}\mathrm{E}[q^{D'}|s,d=A] &= \Pr(t=H|q=1,d=A)(\alpha+\beta+\gamma)+\Pr(t=A|q=1,d=A)\beta\\ &= \beta+\frac{\lambda(\alpha+\gamma)(\alpha+\beta+\gamma)\phi}{\lambda(\alpha+\beta+\gamma)\phi+\beta(1-\lambda)(1-\phi)} < E[q^{D'}|s,d=H].\end{aligned}$$

Proposition 6 (AI detector and the consumer)**.** *The impact of AI detector policy on the consumer is as follows:*

- *(i) If $\bar{c}^D < c < \bar{c}$, AI detector policy does not affect the content quality, and the consumer does not subscribe to the creator's channel independent of the policy.*
- *(ii) If $\bar{c}^{AI} < c < \bar{c}^D$, AI detector increases the average content quality by $\lambda(\alpha+\gamma)$ and increases from 0 to a positive subscription probability and the associated expected quality after subscription increases.*
- *(iii) If $c < \bar{c}^{AI}$, AI detector does not affect the average content quality. If the consumer subscribes when the content detected as human-generated, the expected subscribed quality is higher than the case without AI detector, i.e., $\mathrm{E}[q^{D'}|s,d=H] > \mathrm{E}[q^{AI'}|s]$; if the consumer subscribes when the content detected as AI-generated, the expected subscribed quality is lower than the case without AI detector, i.e., $\mathrm{E}[q^{D'}|s,d=A] < \mathrm{E}[q^{AI'}|s]$. In particular, $\mathrm{E}[q^{D'}|s,d=H]$ is decreasing in ϕ, and $\mathrm{E}[q^{D'}|s,d=A]$ is increasing in ϕ.*

As indicated in Proposition 5, under conditions where AI complementarity γ is low and the detection error rate ϕ is high, the cost of human skill c might surpass the threshold $\bar{c}^D$. In such cases, the high-skill creator is deterred from entering the platform, and consumers do not subscribe to low-skill creators' channels. Consequently, regardless of the detector policy, neither the average quality of the platform nor the subscription rates are affected.

Conversely, when the cost of human skill is moderate ($\bar{c}^{AI} < c < \bar{c}^D$), the high-skill creator is encouraged to enter the platform thanks to the AI detector policy. This policy significantly improves both the average content quality and the expected quality post-subscription by fostering a richer diversity of high-skill content.

In scenarios where the cost of human skill is low ($c < \bar{c}^{AI}$), high-skill creators remain on the platform regardless of the detector policy. In these cases, while the policy does not change the overall mix of creators, thus not affecting the average content quality of the platform, it still plays a crucial role in helping consumers differentiate between creator types. Specifically, if content is detected as high-skill, the consumer anticipates higher quality in the subscribed content than without the detector policy; conversely, content detected as low-skill leads to lower expectations. Therefore, consumers are more likely to subscribe to channels with content detected as high-skill. As the detection error rate ϕ increases, the expected quality for content identified as high-skill also rises, further increasing the consumer's likelihood to subscribe to such creators.

Proposition 7 delineates the comparative impacts of AI watermark versus AI detection policies on the creator and consumer. For simplicity, we define $\hat{\rho}(\phi)$ as follows:

$$\hat{\rho}(\phi)=\frac{(\beta+(\alpha+\gamma)\lambda)(1-\phi)\phi}{(\alpha+\gamma)\lambda(1-\phi)\phi+\beta\left(1-\lambda(1-2\phi)^2-3(1-\phi)\phi\right)}$$

Proposition 7 (AI watermark vs. AI detector)**.** *Assuming any $c < \min\left\{\bar{c}^W, \bar{c}^D\right\}$, the following assertions are valid:*

- *(i) For the high-skill creator, if $\rho > \hat{\rho}(\phi)$, then the AI watermark policy is more beneficial than the AI detector policy, i.e., $U_H^D > U_H^W$. Conversely, if $\rho < \hat{\rho}(\phi)$, the AI detector policy is preferable to the AI watermark policy, i.e., $U_H^D < U_H^W$.*
- *(ii) For the consumer, while the average quality of content remains the same under both policies, the expected quality of content from channels subscribed to varies: if $\rho > \frac{\phi}{1-\phi}$ and the consumer subscribes to channels detected as human-generated, then the expected quality is higher under the AI detector policy than under the AI watermark policy. If not, the expected quality post-subscription is lower under the AI detector policy than under the AI watermarking policy.*

Proposition 7 reveals critical insights for content creation platforms, suggesting that the choice between AI watermark and detection technologies significantly influences the competitive dynamics between high-skill and low-skill creators. For creators, understanding the relationship between the watermark removal rate ρ and the detection error rate ϕ helps to choose the most beneficial platform, optimizing their expected utility. Platform operators must carefully balance these technologies to maintain a fair competitive environment, which is essential for attracting a diverse creator base and ensuring consumer satisfaction. Moreover, the findings underscore the need for regulatory frameworks that support transparency and fairness in AI applications, thereby enhancing consumer trust and fostering a sustainable ecosystem for all stakeholders in the digital content economy.

Concluding Remarks

In conclusion, our study shows that the impact of AI on content creation depends on the cost of human input, the complementarity between AI and human skills, and the presence of AI detection methods. While AI can benefit low-skill creators and enhance high-skill creators' work in some cases, it might also lead high-skill creators to exit the market, reducing overall content quality. AI watermarking and detection services can help mitigate this by allowing consumers to distinguish human-made content and incentivize high-skill creators to stay. The effectiveness of these methods depends on factors like human input cost, AI complementarity, and watermark removal rates. When choosing between watermarking and detection, watermarking might be preferable for high-skill creators if the watermark is difficult to remove, while detection might be better for consumers if the watermark is easily bypassed.

References

Acemoglu, D., Autor, D., Hazell, J., and Restrepo, P. 2022. "Artificial intelligence and jobs: Evidence from online vacancies," *Journal of Labor Economics* (40:S1), pp. S293–S340.

Acemoglu, D., and Restrepo, P. 2018. "Artificial intelligence, automation, and work," in *The economics of artificial intelligence: An agenda*, , University of Chicago Press, pp. 197–236.

Agrawal, A., Gans, J. S., and Goldfarb, A. 2019. "Artificial intelligence: the ambiguous labor market impact of automating prediction," *Journal of Economic Perspectives* (33:2), pp. 31–50.

Amaldoss, W., Du, J., and Shin, W. 2021. "Media Platforms' Content Provision Strategy and Source of Profits," *Marketing Science* (40:3), pp. 527–547.

Autor, D. H., Levy, F., and Murnane, R. J. 2003. "The skill content of recent technological change: An empirical exploration," *The Quarterly journal of economics* (118:4), pp. 1279–1333.

Berman, R., and Katona, Z. 2020. "Curation algorithms and filter bubbles in social networks," *Marketing Science* (39:2), pp. 296–316.

Bhargava, H. K. 2022. "The Creator Economy: Managing Ecosystem Supply, Revenue Sharing, and Platform Design," *Management Science* (68:7), pp. 5233–5251.

Brummer, C. 2024. "Digital Theft and Art," .
URL https://www.artiststrong.com/digital-theft-and-art/

Brynjolfsson, E., and McAfee, A. 2014. *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*, WW Norton & Company.

Callaway, E. 2022. "The entire protein universe': AI predicts shape of nearly every known protein," *Nature* (608:7921), pp. 15–16.

Claussen, J., Kretschmer, T., and Mayrhofer, P. 2013. "The effects of rewarding user engagement: The case of Facebook apps," *Information Systems Research* (24:1), pp. 186–200.

De Cremer, D., and Kasparov, G. 2021. "AI Should Augment Human Intelligence, Not Replace It," *Harvard Business Review* .
URL https://hbr.org/2021/03/ai-should-augment-human-intelligence-not-replace-it

Han, Y., Lappas, T., and Sabnis, G. 2020. "The Importance of Interactions Between Content Characteristics and Creator Characteristics for Studying Virality in Social Media," *Information Systems Research* (31:2), pp. 576–588.

Jain, S., and Qian, K. 2021. "Compensating online content producers: A theoretical analysis," *Management Science* (67:11), pp. 7075–7090.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. 2021. "Highly accurate protein structure prediction with AlphaFold," *Nature* (596:7873), pp. 583–589.

Li, F., and Du, T. 2011. "Who is talking? An ontology-based opinion leader identification framework for word-of-mouth marketing in online social blogs," *Decision Support Systems* (51:1), pp. 190–197.

Mallipeddi, R. R., Janakiraman, R., Kumar, S., and Gupta, S. 2021. "The Effects of Social Media Content Created by Human Brands on Engagement: Evidence from Indian General Election 2014," *Information Systems Research* (32:1), pp. 212–237.

Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., Ko, R., and Sanghvi, S. 2017. "Jobs Lost, Jobs Gained: Workforce Transitions in a Time of Automation," McKinsey Global Institute, available online at: https://www.mckinsey.com/~/media/BAB489A30B724BECB5DEDC41E9BB9FAC.ashx.

McAfee, A., and Brynjolfsson, E. 2017. *Machine, platform, crowd: Harnessing our digital future*, WW Norton & Company.

Qian, K., and Jain, S. 2024. "Digital Content Creation: An Analysis of the Impact of Recommendation Systems," *Management Science* Forthcoming.

Simon, F. M., Altay, S., and Mercier, H. 2023. "Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown," *Harvard Kennedy School Misinformation Review* (4:5).

Singh, N., Jain, M., and Sharma, S. 2013. "A survey of digital watermarking techniques," *International Journal of Modern Communication Technologies and Research* (1:6), p. 265,852.

Susarla, A., Oh, J.-H., and Tan, Y. 2012. "Social networks and the diffusion of user-generated content: Evidence from YouTube," *Information Systems Research* (23:1), pp. 23–41.

Tang, Q., Gu, B., and Whinston, A. B. 2012. "Content contribution for revenue sharing and reputation in social media: A dynamic structural model," *Journal of Management Information Systems* (29:2), pp. 41–76.

The White House 2023. "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence," https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

Toubia, O., and Stephen, A. T. 2013. "Intrinsic vs. image-related utility in social media: Why do people contribute content to Twitter?" *Marketing Science* (32:3), pp. 368–392.