https://pubsonline.informs.org/journal/msom

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 26, No. 1, January–February 2024, pp. 330–349
ISSN 1523-4614 (print), ISSN 1526-5498 (online)

Selecting Cover Images for Restaurant Reviews: AI vs. Wisdom of the Crowd

Warut Khern-am-nuai,[a,*] Hyunji So,[a] Maxime C. Cohen,[a] Yossiri Adulyasak[b]

[a] Desautels Faculty of Management, McGill University, Montreal, Quebec H3A 165, Canada; [b] Department of Logistics and Operations Management, HEC Montreal, Montréal, Québec H3T 2A7, Canada

*Corresponding author

Contact: warut.khern-am-nuai@mcgill.ca, https://orcid.org/0000-0002-1028-1593 (WK-a-n); hyunji.so@mcgill.ca, https://orcid.org/0009-0003-4173-2419 (HS); maxime.cohen@mcgill.ca, https://orcid.org/0000-0002-2474-3875 (MCC); yossiri.adulyasak@hec.ca, https://orcid.org/0000-0002-6996-0742 (YA)

Received: November 3, 2021
Revised: December 19, 2022; April 30, 2023
Accepted: July 30, 2023
Published Online in Articles in Advance: September 7, 2023

https://doi.org/10.1287/msom.2021.0531

Abstract. ***Problem definition*:** Restaurant review platforms, such as Yelp and TripAdvisor, routinely receive large numbers of photos in their review submissions. These photos provide significant value for users who seek to compare restaurants. In this context, the choice of *cover images* (i.e., representative photos of the restaurants) can greatly influence the level of user engagement on the platform. Unfortunately, selecting these images can be time consuming and often requires human intervention. At the same time, it is challenging to develop a systematic approach to assess the effectiveness of the selected images. ***Methodology/results*:** In this paper, we collaborate with a large review platform in Asia to investigate this problem. We discuss two image selection approaches, namely crowd-based and artificial intelligence (AI)-based systems. The AI-based system we use learns complex latent image features, which are further enhanced by transfer learning to overcome the scarcity of labeled data. We collaborate with the platform to deploy our AI-based system through a randomized field experiment to carefully compare both systems. We find that the AI-based system outperforms the crowd-based counterpart and boosts user engagement by 12.43%–16.05% on average. We then conduct empirical analyses on observational data to identify the underlying mechanisms that drive the superior performance of the AI-based system. ***Managerial implications*:** Finally, we infer from our findings that the AI-based system outperforms the crowd-based system for restaurants with (i) a longer tenure on the platform, (ii) a limited number of user-generated photos, (iii) a lower star rating, and (iv) lower user engagement during the crowd-based system.

Funding: The authors acknowledge financial support from the Social Sciences and Humanities Research Council [Grant 430-2020-00106].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/msom.2021.0531.

Keywords: online review platforms • user-generated photos • deep learning • wisdom of the crowd

1. Introduction

User-generated content has become a vital component of consumers' decision making and firms' operations (Kumar et al. 2018). In this context, third-party platforms that collect and share user-generated content have increasingly encouraged contributors to include user-generated photos in their content to satisfy consumers' appetite for rich media. However, the growing amount of user-generated photos poses a significant challenge to the operations of online platforms. More precisely, platforms can receive thousands of user-generated photos for each business entity and cannot display them all. In practice, review platforms need to select only a few photos to serve as *cover images* and display them on important pages, such as the search result page. Although selecting cover images may appear to be a common operational task at first glance, its implications are far from straightforward. First, prior studies have shown that images have a strong influence on user interactions (Zhang et al. 2022). In that regard, an improved cover image selection approach can significantly increase consumer engagement in terms of the number of clicks on the images, which is strongly correlated with the revenue of online platforms that rely on the cost-per-impression (CPM) advertisement model. Second, with the increasing volume of user-generated content routinely supplied to platforms, one of the main challenges that most operations managers face is to handle such user-generated content. Previous literature has shown that the use of manual labor or the crowd to handle user-generated content may not be sustainable and is not scalable (Wang et al. 2019). As such, it is important for online platforms to explore a systematic scalable approach to select cover images that does not rely on manual labor or on the crowd. With many recent success stories on leveraging artificial intelligence (AI) models for business and

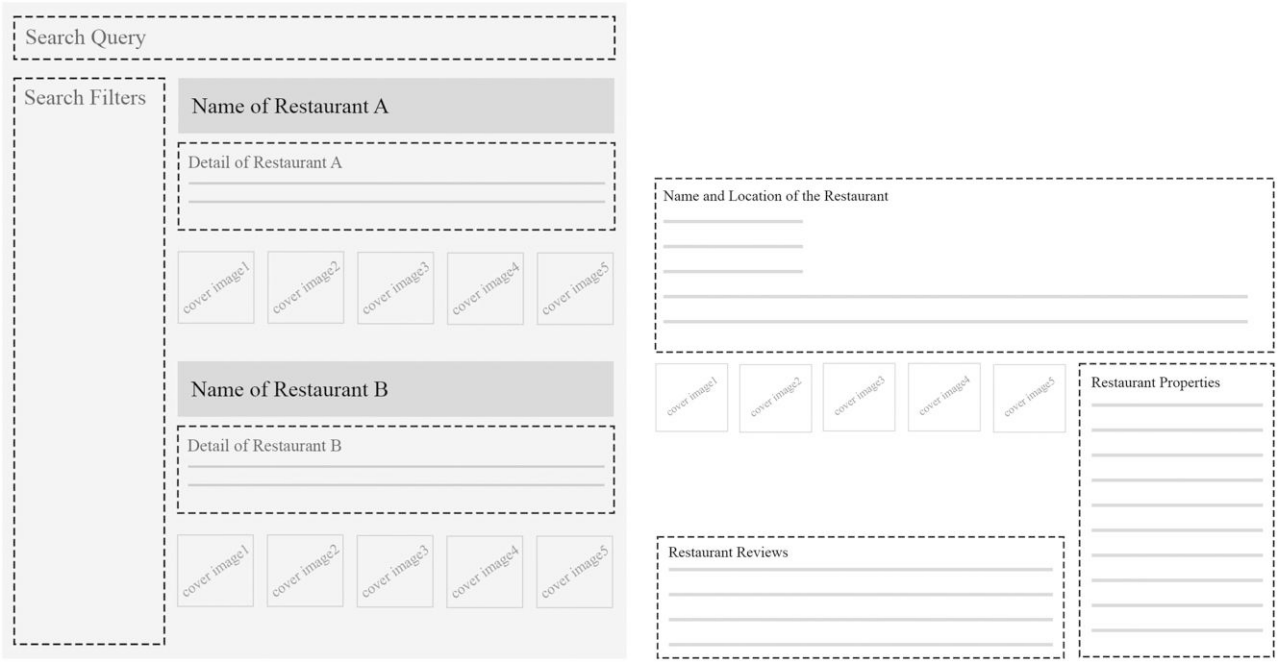
operational decisions (e.g., Cohen 2018, Adulyasak et al. 2023), several platforms have naturally explored the use of an AI-based system to select cover images. However, note that most success stories of AI implementations involve comparing AI with a single (or a few) human agents. Meanwhile, the relative performance between AI and the crowd, which is the empirical context of this paper, is not as straightforward. Specifically, previous studies have argued that AI tends to outperform a human agent because it possesses a higher capability to process information (e.g., Keding 2021). Nevertheless, when the judgement is collectively rendered by a group of individuals (i.e., a crowd), such an advantage is virtually erased (Surowiecki 2004). Especially for tasks that require a subjective assessment or evaluation, prior studies have shown that the crowd can sometimes outperform AI (e.g., Cui 2020). As such, the relative performance of AI versus the crowd in selecting cover images is not ex ante obvious and is an open empirical question. Inspired by this gap in the literature, we aim to address the following research question.

Can we develop a practical AI model that outperforms the crowd in selecting cover images that stimulate user interactions on an online review platform?

To answer this question, we collaborate with a large restaurant review platform in Asia. The cover images of restaurants on the platform were initially selected by a crowd-based system (see Figure 1). However, the platform was aware that such a conventional approach could have shortcomings in terms of sustainability. We thus partner with the platform to develop an alternative AI-based system that exploits a transfer learning approach based on a state-of-the-art computer vision model to select cover images. It is worth noting that AI and the crowd tend to select different photos as cover images (detailed discussions can be found in Section 5.3.1). In our study, we present the comparison between the crowd-based system and the AI-based system from an operational perspective. Namely, from the platform's perspective, the AI-based system and the crowd-based system are two possible alternatives. As such, our study provides insights on the relative performance of both systems, which assist the platform's operations manager who faces an operational decision of choosing between these systems.

Our study first validates the AI-based system via both an internal validation (out-of-sample testing) and an external validation (randomized field experiment). Overall, our results demonstrate that the AI-based system significantly outperforms the crowd-based system in terms of its ability to select cover images that stimulate user interactions (captured by the number of clicks generated by users on the platform). Following these favorable results, the AI-based system was fully deployed on the platform in 2019. Using a comprehensive data set including observations both before and after the implementation, we further validate that the AI-based system indeed outperforms the crowd-based system in stimulating user interactions on the platform. In addition, we examine the heterogeneity in the performance of the AI-based system with respect to several restaurant characteristics. We also conduct image-level analyses to identify the potential underlying mechanisms that lead to the difference in performance between the AI- and crowd-based systems. Interestingly, we find that AI outperforms the crowd in selecting images with superior intrinsic characteristics (e.g., higher brightness and saturation), which are the

Figure 1. The Position of Cover Images in the Search Result Page (Left Panel) and Restaurant Landing Page (Right Panel)



factors that positively affect user interactions. Meanwhile, the crowd is more likely to be influenced by nonimage characteristics (e.g., the number of followers of the user who uploaded the image), which do not positively impact user interactions. Lastly, we perform post hoc analyses to verify that the increase in the number of clicks induced by the adoption of the AI-based cover image system does not cannibalize other important user participation metrics, such as photo votes and photo uploads.

Insights from our research contribute to the operations management (OM) literature both for researchers and for practitioners as follows. First, operations managers of online platforms usually face an increasing amount of user-generated content, and handling this content has become one of the prominent operational issues because of the required amount of administrative efforts (Wang et al. 2019). In that regard, our work empirically demonstrates that the use of an AI model can help reduce manual labor in evaluating user-generated content, hence improving operational efficiency. Note that the focal AI model in our study delivers an improved number of clicks, which is one of the main key performance metrics from the platform's perspective. As such, our work contributes to the recent OM literature on the operational value of artificial intelligence and user-generated content (e.g., Cui et al. 2022, Mithas et al. 2022). Second, our paper shows that the use of AI can help improve matches between customers and service providers (in our case, restaurants) to determine desired dining places. Because service performance has been identified to be largely a function of fit between customers and operational outcomes, our work contributes to the literature on consumer matching in the context of service operations management (e.g., Manshadi et al. 2022, Aouad and Saban 2023).

2. Related Literature

We next review related literature. First, we review the literature on the wisdom of the crowds and on the effectiveness of user-generated content. Second, we survey several related studies that leverage deep learning methods in management research. Third, we discuss the recent research stream on field experiments in online platforms. Lastly, we explore past studies that examine human-algorithm connections.

2.1. Wisdom of the Crowds and User-Generated Content

Wisdom of the crowds refers to the concept of collective opinions being better than an opinion of an individual (Surowiecki 2004). It has been extensively used in management and economics to explain the power of aggregated decisions or opinions. Prior studies generally used the concept of wisdom of the crowds to justify the underlying influence of user-generated content (e.g., Khern-am-nuai et al. 2023a). For example, Liu and Karahanna (2017) argue that online reviews that represent aggregated opinions are helpful in assisting consumer purchase decisions. Kittur et al. (2007) demonstrate that the success of Wikipedia can be attributed to wisdom of the crowds. Similarly, many firms leverage the wisdom of the crowds as the core concept to solve challenges or gather ideas. For instance, Koh (2019) studies the influence that a firm can have on the behavior of the crowds in ideation contests.

In the context of this paper, the platform leverages wisdom of the crowds to select cover images (i.e., selected representative images from user-generated photos) by relying on platform users' votes. User-generated photos refer to the photos taken and attached by the users to provide additional information in their review. Several recent studies have shown that this type of visual content has strong implications on the platform's welfare. For example, Cheng and Ho (2015) analyze 983 online restaurant reviews and show that reviews with more user-generated photos are perceived to be more useful by review readers. Follow-up studies based on a large-scale data set also find consistent results. For instance, Yu et al. (2023) use consumer reviews from an online restaurant review platform and find that reviews with more photos tend to receive more helpfulness votes, indicating that user-generated photos are perceived to be valuable by review readers. Overgoor et al. (2020) use a computer vision model to uncover how the visual information of hotel images affects the decision-making process in an online hotel booking system.

2.2. Deep Learning in Management Research

With the increasing importance of unstructured data such as text, voice, and images, AI models that can systematically handle this type of content have been widely adopted (Zhang et al. 2018). Many studies have used deep learning-based models to perform text mining tasks. For instance, deep learning is used to uncover latent characteristics and traits (e.g., Adamopoulos et al. 2018). It is also used for other text mining tasks, such as sentiment analysis (Kokkodis and Lappas 2020). Alternatively, several research projects have used deep learning as a core design element in a multitude of settings. For example, Shin et al. (2020) propose a visual data analytics framework based on deep learning and apply it to social media case studies. Guo et al. (2018) develop a deep learning model that analyzes users' check-in behavior and integrates geographical and social influence to generate personalized point-of-interest recommendation. Lastly, Zhang et al. (2022) leverage a convolutional neural network (CNN) to quantify the aesthetic quality of images and use the predicted image quality as an explanatory variable to showcase the economic value of verified photos on the Airbnb platform. The authors used three categories of low-level attributes: components-composition, color, and figure-ground relationship as

variables. Instead of relying on low-level attributes, our study focuses on using a deep learning-based image assessment model, which relies on a complex latent representation of image features, to select cover images. We leverage a randomized field experiment to validate the effectiveness of our AI-based system in stimulating user interactions. We then exploit observational data to draw insights on the underlying mechanisms that drive the differences between the performance of the AI- and crowd-based systems.

2.3. Field Experiments in Online Platforms

Field experiments are considered the gold standard for causal inference and for comparing two alternatives. Online platforms are the perfect medium to support and deploy field experiments for validating business intuitions and assessing the effectiveness of different strategies and technologies (Kohavi and Thomke 2017). Many researchers have recently used field experiments in the context of online platforms in various scenarios (e.g., Gallino and Moreno 2018; Huang et al. 2019; Sun et al. 2021; Cohen et al. 2022, 2023; Cui et al. 2022). In this paper, we perform a randomized experiment to assess the impact of an AI-based system that selects cover images for a restaurant review platform. More generally, this paper is related to the growing stream of empirical studies in the context of online platforms (see, e.g., Xu et al. 2021, 2023) and in particular, on the restaurant industry (see, e.g., Tan and Netessine 2020, Feldman et al. 2023).

2.4. Human-Algorithm Connections

With the increasing adoption of AI in operations, several recent studies have investigated the complementarity between humans and algorithms. For example, Cui et al. (2022) examine how AI creates value in procurement when the price inquiry process is automated in an online business-to-business platform. This study demonstrates that suppliers tend to vary their wholesale price when AI replaces human tasks in price inquiries, and the automation is only valuable when carefully implemented. In addition, Bai et al. (2022) investigate how an automated process leveraging AI (versus humans) affects recipients' task productivity and their perception regarding the fairness of the task assignments. Their findings suggest that people perceive AI task assignments as fairer than human-based traditional systems, which in turn, improves their productivity. Our paper complements prior studies in this area that examine the human-algorithm connection in several operations management contexts, such as repetitive tasks (Spring et al. 2022), healthcare (Kyung and Kwon 2022), and collaborative decision making (Fügener et al. 2022).

3. Research Context

This section describes the general context of our research. We first discuss the concept of cover images, which is our main focus of this paper. We then describe the current crowd-based image system, where cover images are selected by platform users.

3.1. Platform Partner and Cover Images

With the growing importance of user-generated photos, our review platform partner decided (back in 2014) to offer reviewers the option to attach self-taken photos of food and beverages for restaurant reviews. As expected, this feature was positively received by platform users, allowing restaurants on the platform to reach thousands of images submitted by reviewers (more detailed descriptive statistics are presented in the sequel). Naturally, such a high volume of user-generated photos warrants that the platform develops a proper management plan. An important aspect related to image management is to decide the order in which these images are displayed on the platform, especially on important pages. As such, the platform uses a cover image system, where five user-generated photos are selected as representative images for each restaurant. These cover images are displayed on two of the most important pages of the platform: the search result page and the restaurant landing page, as illustrated in Figure 1. Note that both pages' details, except for the cover images, are blurred to preserve the anonymity of the platform.

Evidently, deciding which photos to use as cover images is critical both to the platform and to the restaurants as they serve as one of the main appeals to attract users' attention on the most frequently visited pages. Thus, the challenge from the platform's perspective is to carefully choose and display images that can bring the highest level of user engagement. In this regard, the platform initially developed and deployed a system to select cover images based on the crowd's opinion (i.e., evaluations from users). We next discuss this common cover image system.

3.2. Traditional Crowd-Based Cover Image System

As discussed, a common approach to selecting cover images is to rely on the "wisdom of the crowd" concept. Specifically, this system selects cover images based on evaluations from the crowd. Although there is no prior research that formally shows the effectiveness of this system in the specific context of user-generated photos, previous studies in related areas have shown that the crowd is consistently effective in quality evaluation as long as generic conditions, such as the people making these choices being reasonably diverse and independent, are met (Surowiecki 2004). For instance, many platforms have relied on user evaluations to determine the quality of user-generated reviews (Yu et al. 2023). As such, review helpfulness votes, which are cast by platform users, have been widely adopted by researchers as a proxy to measure review quality (e.g., Otterbacher 2009). Therefore, by extrapolation, it is reasonable to believe that a cover image system based on the crowd's evaluations

would perform well for selecting images that stimulate user engagement with the platform.

With such a positive theoretical prediction, the platform implemented the crowd-based cover image system in mid-2014 right after allowing reviewers to attach user-generated photos in their reviews. Specifically, the guideline on the platform is for its users to select "images that they deem helpful for fellow members in determining the restaurants to dine in." From the platform's perspective, the purpose of this system is to identify images that can generate the best user interactions (i.e., users who observe pertinent images would be inclined to further explore the corresponding restaurant). It was thus decided to use the five user-generated photos that receive the highest number of votes for each restaurant as the cover images for that restaurant. In this system, each registered platform user can vote for each image at most once. When a user accesses the photo page of a restaurant, all the photos related to the restaurant are shown to the user. By default, these photos are sorted by recency, and the platform does not use any algorithm to prioritize displaying specific photos to certain users. The platform also implements a monitoring system to detect fraudulent activities, such as creating multiple accounts to vote. Nonuser-generated photos (e.g., photos professionally taken by the restaurants) are also strictly prohibited by the platform. This feature has been well received by users, as they have been actively participating in the voting system. Specifically, cover images selected by this system receive in general more than a thousand votes.

Even though the crowd-based cover image system has performed reasonably well since its implementation, the platform managers have three primary concerns regarding this design. First, the core function of this design critically depends on user participation, which is outside the platform's control. Second, it remains unclear whether the cover photos, which are conventionally selected based on users' votes, are particularly effective in stimulating user interactions. Third, the system requires increasing amounts of intervention from the platform. For instance, the platform had to hire a dedicated support team to review potential malicious photo voting behavior, which is potentially not scalable as the number of restaurants and users continue to increase. Ultimately, the sustainability of this design has become one of the platform's main concerns. Thus, the platform expressed an interest in partnering with us to develop an alternative cover image system that can operate with fewer interactions between users and platform administrators (and hence, reduce operating and maintenance costs). We next describe the design and validation of our alternative cover image system that leverages a state-of-the-art AI model.

4. The AI-Based Cover Image System

In this section, we discuss the AI-based system, where cover images are selected by a deep learning model. From a technical standpoint, this is an image assessment system where the platform relies on an automated process to identify images that would be perceived most appealing by users and select them as cover images. In that regard, the platform was facing the dilemma of either adopting an already existing system or developing a new customized one. For that reason, we next survey the existing computer vision models that could satisfy the platform's needs.

4.1. Survey of Existing Models

We review the computer vision models that provide image assessment capabilities. We begin by examining image aesthetic quantification models that use a supervised learning approach. We then survey a well-known generic image assessment model that is widely adopted in various applications.

4.1.1. Supervised-Based Image Aesthetic Quantification Models. The first type of computer vision models that could satisfy the platform's requirement in using an AI model to select cover images is image aesthetic quantification models. These models use supervised machine learning methods to quantify the aesthetic of images. Because previous studies have shown that high aesthetic images tend to attract user attention (e.g., Park et al. 2017), this type of model could be used to select cover images based on an aesthetic score generated by the model. For instance, Miller (2016) proposes a supervised image aesthetic quantification model that utilizes intrinsic image characteristics as predictors, including low-level features (e.g., lightness, colorfulness, color harmony) and high-level features (e.g., object alignment, depth of field, golden ratio). Using these features, the target variable of the predictive task is set to be the image aesthetic score, which is manually assigned by humans. Such a predictive task, however, has two notable limitations. First, there are virtually no standard guidelines for the feature selection process, which makes the feature extraction and selection highly subjective. For example, the features used in Miller (2016) are different when compared with other studies, such as Lou and Yang (2018) and Zhang et al. (2022). Second and more importantly, these supervised models require a large training data set for which the data need to be labeled. In other words, if the purpose of the model is to quantify the aesthetic of images, the photos in the training set are required to have their aesthetic score calculated beforehand. In the context of this study, the photos available on the platform do not have aesthetic scores, and it would be impractical to manually assign such a score to all photos. Because of these limitations, it is difficult for the review platform to directly adopt existing image aesthetic quantification models. Nevertheless, because prior literature has shown that image aesthetic tends to successfully attract user attention, the latent features of images with high aesthetic may

benefit our image assessment model in increasing user engagement. As we discuss in Section 4.2, the design of our model takes advantage of latent features of a pretrained image aesthetic quantification model combined with transfer learning. We next examine a state-of-the-art computer vision model to identify a framework that can be leveraged for our model development.

4.1.2. Generic Image Assessment Model. As discussed, we decided to opt for a solution design that leverages an image quality assessment model in the pretraining phase (via transfer learning). We then need to evaluate the applicability of an existing generic image assessment model in terms of its ability to assess and select images in our data set. In that regard, the model of interest is the Neural Image Assessment (NIMA) model (Talebi and Milanfar 2018). It is considered one of the most popular computer vision models for image assessment tasks.

NIMA is an image assessment model that was built to quantify the technical quality (e.g., pixel-level degradation) and aesthetic (semantic-level characteristics such as emotions and beauty) of images. Technically, NIMA utilizes a deep CNN to predict the ratings of a typical user based on technical and aesthetic criteria. The model is flexible on classifier architecture as it can use different classifiers, such as VGG16 (Simonyan and Zisserman 2014), Inception-v2 (Szegedy et al. 2016), and MobileNet (Howard et al. 2017). Section A in the online appendix summarizes the performance of the NIMA model for different choices of classifier along with the performance of other image assessment models that are closely related to NIMA.

Once the classifier is chosen, the model is first trained on the ImageNet data set, which is a large-scale publicly available image classification data set (Krizhevsky et al. 2012). Although this data set was originally provided for object detection and image classification tasks, the scale and variety of the images in this data set make it suitable to pretrain the classifier in order to allow the neural network to learn and extract common image features, which will be later used for image assessment. Following this step, the model is fine-tuned for the purpose of technical quality and aesthetic quantification by using it on three image data sets: the AVA database, Tampere Image Database 2013 (Ponomarenko et al. 2013), and the LIVE in the Wild Image Quality Challenge Database (Ghadiyaram and Bovik 2015). The AVA database is used for aesthetic evaluation, whereas the latter two data sets are used for technical quality evaluation.

Using NIMA overcomes the technical limitations of supervised image aesthetic quantification models discussed in Section 4.1.1. More precisely, NIMA relies on a CNN for feature selection to alleviate the subjectivity issue of the process. In addition, NIMA trains the model using multiple publicly available image data sets, so that the platform no longer needs to go through the tedious process of labeling its own photos. Nevertheless, it is worth noting that NIMA was built to be a *generic* image assessment model. In other words, the standard NIMA model was developed to predict the ratings of a typical user and was trained on three different data sets. As such, the standard NIMA model performs reasonably well for classifying the aesthetic quality of generic images (as presented in Table 1 as well as in Talebi and Milanfar 2018). In our setting, however, in which the images are exclusively food and beverages, the model did not perform well because it utilizes a large number of features to capture the diverse aspects of image semantics. During our internal validation (detailed discussions can be found in Section 4.3.1), we also included the standard NIMA model as a benchmark, and we observe that the performance of the standard NIMA model is significantly lower relative to our proposed model. Because of these reasons together with our partner platform, we have decided not to adopt the standard NIMA model. Instead, we opt to use NIMA as a base model and leverage a transfer learning approach to further improve the model performance.

Table 1. Performance of Crowd-Based vs. AI-Based Systems (Metric: Number of Users Who Click on Cover Images on the Search Result Page)

	Crowd-based system	AI-based system
Total users	12,082	7,668
Primary measure	2,553	1,991
Click ratio, %	21.13	25.97
Differences	4.84 (22.91% ▲)	
	z-score: 3.3111, p-value < 0.001	

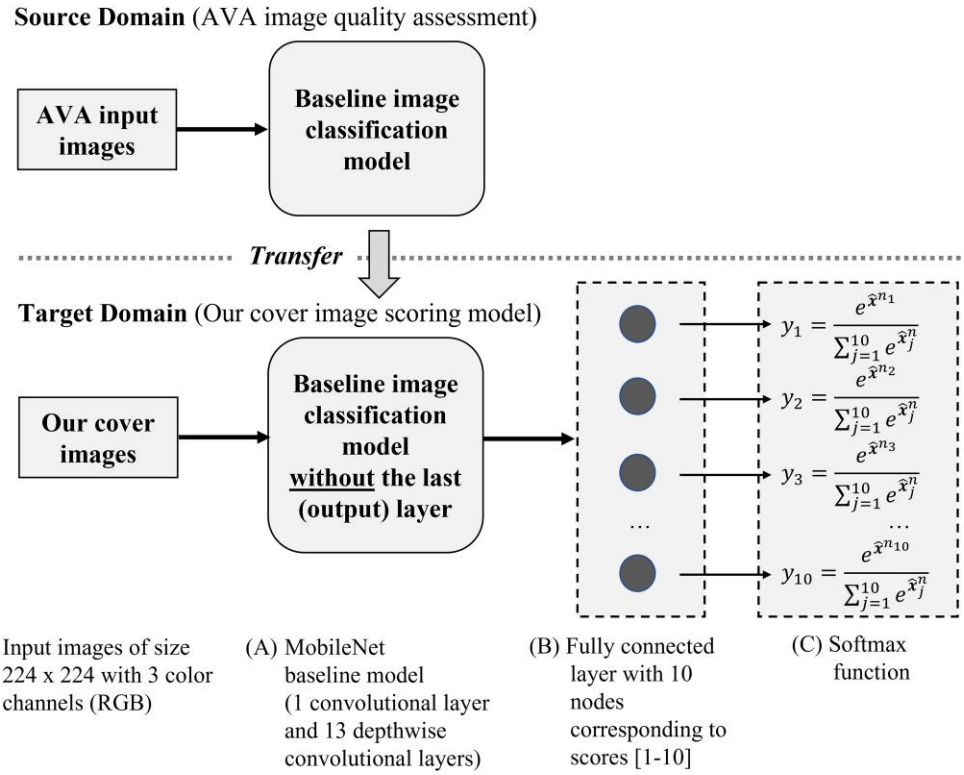
4.2. Model Development

Transfer learning is an emerging technique, in which a machine learning model that was developed for a specific task (called the base model) is reused for another task (e.g., Pan and Yang 2009). This approach is particularly popular in deep learning models, especially in computer vision and natural language processing, because it can leverage the learned feature representation of the pretrained (base) model and successfully apply it to a similar task. In addition, it can substantially reduce the computing time and resources required to retrain the model from scratch. In the transfer learning process, the first model is trained using the data of the original or source domain. This trained model is then leveraged in a different domain (called target domain), which is typically similar to the original domain. In our case, the model in the source domain is a generic image quality assessment model, whereas the model required for the target domain is the cover image scoring model. Because the source and target domains are different, part of the model previously trained must be modified.

To this end, we leveraged the *network-based deep transfer learning approach*, which is widely adopted in deep learning for computer vision and machine translation (Tan et al. 2018). More specifically, rather than training the entire model from scratch for the target domain or data set, this approach reuses parts of the deep learning network (i.e., its structure and its initial parameters) that is pretrained using the source data set as the initial network of the target data set. In deep learning models, the last few layers that compute target variables are typically retrained using the data of the target domain in the transfer learning process. This method is very useful in practice, especially when the amount of data in the target domain is limited. The training process applied to our setting is illustrated in Figure 2, and the detailed description of the deep learning model is provided in Online Appendix B. We note that the structure of the original model except the last layer remains the same when performing transfer learning. Consequently, this process does not increase the computational complexity in the training process because of the fact that it fine-tunes only the last layer of the neural network by using the gradient descent algorithm (Tan et al. 2018). More importantly, because the training process is done offline prior to deployment and the deep learning model scales linearly with the data size (Iandola et al. 2016), this training process does not affect the performance of the prediction process in the pipeline, which is performed using the best-performing NIMA model combined with transfer learning.

We begin our model development by customizing the NIMA model to make it more suitable for our image assessment task. First, we choose a classifier (component (A) in Figure 2). As previously mentioned, NIMA is flexible in terms of the classifier architecture. In our case, we select MobileNets as the primary classifier architecture because of its efficiency.[1] Recall that our model is used essentially for every user-generated photo available on the platform. Consequently, the ability to generate results in a timely manner is highly desirable. Second, in the pretraining phase to train the model for the source domain, we train the NIMA model based on the AVA data set. The use of such a data set in the transfer learning process is motivated by the fact that the label available in the data set (i.e., image aesthetic) is closely related to the objective of our image assessment task (i.e., the ability to stimulate user engagement), as noted in prior studies (e.g., Park et al. 2017). Thus, training the model by using the AVA data set would not only be efficient in terms of resources but would also improve the model accuracy. Third, generally, the output layer of a neural network model that uses the MobileNets classifier would be a fully connected layer with 1,000 nodes. This follows from the fact that the default objective of MobileNets is object recognition. The 1,000 nodes in the output layer represent 1,000 possible objects, and the value associated

Figure 2. Illustration of the Training Process with Transfer Learning



Note. RGB, red, green, and blue.

with each node is the probability that the image contains each object. In our setting, however, the objective is to assess the image in terms of its ability to boost user engagement. Therefore, in the target domain (cover image scoring), we adapt the output layer to contain only 10 nodes, and the value of each node represents the probability that the image would be assigned a score ranging between 1 and 10.[2] This is done by replacing the last layer in the original baseline model used in the source domain by a fully connected layer with 10 nodes (component (B) in Figure 2). The output vector from each node in the last layer is denoted by $\bar{x}_i$, where i is the output associated with score i and then converted into the probability value of score i (denoted y_i) using the softmax function as shown in component (C) in Figure 2. The retraining process of transfer learning is performed in two phases: *freeze training* and *full training*. In the freeze training phase, the weights in each node except for the nodes in the last layer (component B) are frozen (i.e., not updated). Indeed, the early convolution layers (i.e., layers close to the input layer) are mainly used for feature extraction, whereas the late convolution layers (i.e., layers close to the output layer) are primarily used for classification. The purpose of the freeze training phase is to allow the model to focus more on adapting to the transformed features that are extracted from food and beverage images using the learned representations from the source domain. Subsequently, we perform the full training phase, which allows the weights of all nodes in processing layers to be updated to fully optimize the performance of the model based on the data from the target domain. Last, we define the loss function (i.e., the function that calculates the difference between the model output and its true value). The model then uses these values to determine the optimal weights in the neural network. We choose the Earth Mover's Distance (EMD) as our model's loss function given that prior studies have shown that EMD works well for neural network models, where all the nodes in the output layer are related to each other (e.g., Rubner et al. 1998).

At this point, the standard NIMA model is customized for our image assessment task. Next, we apply a transfer learning process to this customized NIMA model. We highlight that the performance of the final model will heavily depend on the training data set. In this case, the ideal training data set would contain exclusively labeled food and beverage images. Unfortunately, the images in our platform data set are not labeled. Therefore, we construct such a training data set by randomly selecting 1,200 images from our data set and utilize skilled human agents to label each image.[3] We note that the labels of the images performed by human agents in this process would become the "ground truth" in the machine learning model. As such, this procedure could have led to a machine learning bias had the labeling process not been properly executed. For example, consider the case where there is only one skilled human agent who labels the images. In this case, the machine learning model would essentially capture the agent's preferences rather than the intrinsic appeal of the images. To alleviate the possibility of such a bias, we rely on a large pool of platform employees to serve as agents. We then randomly assign each image to five different agents and ask each agent to rate each image using a scale between 1 and 10 based on the same guideline used for platform users (i.e., the helpfulness for platform users in determining the restaurants to dine in). We ultimately use the average of the five scores as the appealing score for training purposes. In this way, we ensure that our model captures the overall appeal of images rather than specific preferences of an agent. To evaluate this process, we use 1,000 images (of the 1,200 images labeled manually by the agents) for training, and the 200 remaining images are kept for an internal performance evaluation.

4.3. Model Validation

4.3.1. Internal Validation. We first evaluate the performance of our model internally. In other words, we aim to validate the model performance based on the holdout data set (composed of 200 out-of-sample images). Specifically, the effectiveness of our model is measured by how the rankings of images, based on the scores from our model, are relevant when compared with human judgments. Recall that when the image assessment model is actually implemented, it has to display images in a specific order (e.g., the first image often leaves the strongest impression to users). Thus, we consider a sorting-based evaluation to reflect this aspect. As such, we adapt the normalized discounted cumulative gain (nDCG) (Järvelin and Kekäläinen 2000) as the primary criterion for our model evaluation (see Section C in the online appendix for more details regarding nDCG). Essentially, the relevance aspect of an image is measured by the score that is predetermined by the judgment of human agents. The cumulative value is then discounted so that if a more relevant image is ranked lower, the cumulative value is penalized. Lastly, the output is normalized by the maximum value (obtained from the perfect rankings using the human agents' judgment), so that the resulting output ranges between zero and one, where zero corresponds to the worst ranking and one indicates a perfect ranking. This measure is often used to evaluate the effectiveness of search-engine algorithms in terms of returning more relevant search results (e.g., Wang et al. 2013). In our context, nDCG would provide a meaningful evaluation of our model in terms of its sortability based on its ability to stimulate user interactions.

In the evaluation process, we first split the 200 testing images into 20 groups, each with 10 images. We attempt to group similar images together based on their characteristics (e.g., similar food type, ethnicity) to ensure relevant comparisons. Our objective is to use our

model to quantify its ability to stimulate user interactions and rank the images according to their scores. We then compare our results with the rank of the scores generated by skilled human agents, as described in Section 4.2. We specify the parameter p to be 10 in nDCG, which represents the maximum position measured (i.e., we measure the performance of our model on every position in the ranked order because each group has up to 10 positions). We execute our model on all 20 groups of images and use the average nDCG score as the evaluation measure of the model. Our best model produces an nDCG score of 0.9219, which is distinctively high relative to prior studies that utilize nDCG as the main measure (e.g., Wang et al. 2013). This result suggests that our model performs exceptionally well when compared with scores generated by human agents.

We highlight that the validation uses human agents' judgment as the ground truth. In practice, however, our desired goal is to stimulate user interactions. Although these two objectives are highly correlated, they are not the same. In other words, maximizing the nDCG score does not necessarily yield the maximal level of user interaction. Nevertheless, this internal validation allows us to gain insights on the AI-based system in terms of its performance and reliability prior to evaluating it externally (i.e., via a randomized field experiment), as we discuss next.

4.3.2. External Validation. Our main goal is to compare the performance of the crowd-based and AI-based cover image systems in terms of their ability to stimulate user interactions with the platform. The gold standard for making such a comparison is to run a randomized field experiment. Given the potential benefits of our AI-based system, the platform partner agreed to perform a live test, which we discuss next.

The randomized field experiment was conducted in early 2019 (the exact dates are omitted because of our nondisclosure agreement). During the experimental period, a small percentage of users who visited the platform were randomly redirected to an experimental system[4] (we cannot reveal the exact percentage, but it corresponds to more than 50,000 users and hence, represents a large-enough sample to make statistical inference). In practice, the platform relies on a third-party provider to run this type of experiment. First, the provider duplicates the platform system into two entities. The experimental system and the platform system are identical (except for the cover image system in use), and users are not aware of the existence of the experiment (i.e., they do not know whether they are on the platform system or on the experimental system).

For users assigned to the experimental system, around 62.5% of them observed cover images selected by the crowd-based system (the number 62.5% was decided based on several discussions with the platform executives). Meanwhile, the rest of the users were directed to the experimental system and observed cover images selected by the AI-based system. The randomization was performed using a common approach in the literature, where the session identification of each visit is used to assign the treatment condition (Lee and Hosanagar 2019, Khern-am-nuai et al. 2023b).[5] We note that, for users who were directed to the experimental system, only fresh visits to the restaurants were included in the data collection (e.g., if user A visited restaurant X before the start of the experiment and visited restaurant X again during the experiment, such a visit was excluded from the data collection). Importantly, to avoid potential violations in terms of data privacy regulations in the country where the platform operates, we only collected aggregate-level data as opposed to granular behavior data (i.e., data tied to a specific user or restaurant). More precisely, when using a third-party provider to facilitate the experimentation process, our platform partner cannot collect any user-related data (because it cannot explicitly ask for user consent). Instead, we rely on aggregate-level data that still allow us to compare the performance of the AI-based and crowd-based systems. Later, in Section 5, we show how we can collect comprehensive user-based data and thus, conduct several refined analyses to draw additional managerial insights.

Recall that cover images are displayed on two pages: the search result page and the restaurant landing page. Thus, we evaluate the performance of the cover image systems on these two pages separately. The platform's primary revenue stream is advertisement, and the platform uses the CPM scheme. As such, the platform's primary interest for this experiment is whether the AI-based cover image system leads to an increase in the number of users who click on one of the cover images. As discussed, this follows from the fact that each click generates page views, which directly translate into platform revenue.

4.3.2.1. Search Result Page. The first page that displays cover images is the search result page, which shows a list of restaurants that match the search query (see the left panel in Figure 1). In addition to the primary performance measure (i.e., the number of users who click on one of the cover images), we also construct a secondary performance measure, which is the number of users who click on the restaurants. In total, there were 19,750 users who were directed to the experimental system during the experimental period, searched for the restaurants, and reached the search result page. Among these users, 12,082 observed the crowd-based cover image system, whereas 7,668 observed the AI-based cover image system (i.e., 37.5% of users were exposed to the AI-based system as initially agreed upon with the platform). The experiment naturally recorded only unique clicks (i.e., if a user clicked on the same restaurant or on the same photo multiple times, only the first click was recorded). Also, the same user could click on both restaurants and photos.

Tables 1 and 2 report the results of the field experiment on the search result page for the primary and secondary measures, respectively. Interestingly, the AI-based system outperforms the crowd-based system in stimulating user interactions by a sizable amount. Specifically, the ratio of users who clicked on cover images under the AI-based system is 22.91% higher relative to the crowd-based system. In addition, the ratio of users who clicked on the restaurant under the AI-based system is 11.76% higher relative to the crowd-based system. The two-sample z-test for independent proportions shows that the differences in both cases are statistically significant with a p-value below 0.001.

4.3.2.2. Restaurant Landing Page.

The second page that displays cover images is the restaurant landing page, which is the main page of the restaurants on the platform (see the right panel in Figure 1). In this case, in addition to the primary measure, we construct a secondary performance measure that captures the number of users who interacted with the restaurant.[6] Similar to the experiment on the search result page, only unique clicks (and interactions) were recorded, and both interactions with cover images and interactions with the restaurant were recorded. In total, there were 30,373 users who were redirected to the experimental system during the experimental period and reached the restaurant landing page. Among these, 18,842 users observed the crowd-based cover image system, whereas 11,531 users observed the AI-based cover image system. Note that, in this experiment, there was no user that was included in both the experiment on the search result page and the experiment on the restaurant landing page. In other words, the treated users who observe cover images selected by the AI-based system on the restaurant landing page will never be treated on the search result page and vice versa. In addition, the users who were included in the experiment on the restaurant landing page reached the restaurant landing page directly and never observed cover images of the corresponding restaurant on the search result page prior to (or during) the experiment. For example, they could access the restaurant landing page through the restaurant website or the social media page, via an external search engine (e.g., Google search), or through newspapers or online articles in which the link to the restaurant landing page is included.

Table 2. Performance of Crowd-Based vs. AI-Based Systems (Metric: Number of Users Who Click on Restaurants on the Search Result Page)

	Crowd-based system	AI-based system
Total users	12,082	7,668
Secondary measure	6,055	4,295
Click ratio, %	50.12	56.01
Differences	5.89 (11.76% ▲)	
	z-score: 4.0383, p-value < 0.001	

Table 3. Performance of Crowd-Based vs. AI-Based Systems (Metric: Number of Users Who Click on Cover Images on the Restaurant Landing Page)

	Crowd-based system	AI-based system
Total users	18,842	11,531
Primary measure	5,783	3,947
Click ratio, %	30.69	34.23
Differences	3.54 (11.53% ▲)	
	z-score: 2.9918, p-value = 0.001	

Tables 3 and 4 report the results of the field experiment on the restaurant landing page for the primary and secondary measures, respectively. Similar to the results on the search result page, the AI-based system significantly outperforms the crowd-based system in terms of user interactions. Specifically, the ratio of users who clicked on the cover images under the AI-based system is 11.53% higher relative to the crowd-based system. In addition, the ratio of users who interacted with the restaurant under the AI-based system is 16.09% higher relative to the crowd-based system. The two-sample z-test for independent proportions shows that the differences in both cases are statistically significant with a p-value of 0.001 and a p-value of <0.001, respectively.

The results from the randomized experiment provide promising evidence that AI outperforms wisdom of the crowds at selecting images to stimulate user interactions. The increase in user interactions that the AI-based cover image system generates is significant both statistically and economically. At the same time, these results have a notable limitation. Specifically, the data are collected and analyzed at the aggregate level only, and the characteristics of each user cannot be observed. Hence, we complement these findings by leveraging observational data that we obtain from the platform-wide implementation of the AI-based system in 2019. Because this is no longer an experiment (and does not involve a third-party provider), we will now be able to collect granular data and conduct more refined analyses. More precisely, in the next section, we empirically examine the performance of the AI-based system after its full deployment on the platform.[7] In addition, we identify the heterogeneity of the performance benefit of the AI-based system and uncover potential underlying mechanisms that explain why AI outperforms the crowd in selecting cover images that stimulate user interactions. Ultimately, our analyses will allow us to refine our managerial insights by understanding for which type of restaurants the AI-based system performs best.

5. Implementation Results

Results from our randomized field experiment showed that the AI-based system outperforms the crowd-based system in selecting cover images that stimulate user interaction. To expand our aggregate-level findings from

Table 4. Performance of Crowd-Based vs. AI-Based Systems (Metric: Number of Users Who Interact with Restaurants on the Restaurant Landing Page)

	Crowd-based system	AI-based system
Total users	18,842	11,531
Secondary measure	6,275	4,458
Action ratio, %	33.30	38.66
Differences	5.36 (16.09% ▲)	
	z-score: 4.5314, p-value < 0.001	

the experiment, we empirically examine the performance of the AI-based system after its deployment on the entire platform. We have access to the platform's observational clickstream data from both before and after the AI-based system was fully deployed in 2019. More precisely, for each restaurant, we have the data for a period of 12 weeks before and 12 weeks after the implementation date (i.e., the platform-wide deployment of the AI-based system). By collaborating with the platform, we randomly select 4,000 restaurants, representing approximately 4% of the restaurant population. We validate that the sample represents the population of the restaurants in Section E.1 in the online appendix.

We preprocess our data to remove the restaurants that are not applicable to our analysis. For example, we naturally exclude restaurants that are permanently closed before the implementation of the AI-based system (such restaurants are likely to have no updated images and much smaller crowd participation).[8] In addition, given that the cover image system selects five images, we remove the restaurants with five or fewer user-generated photos before the implementation of the AI-based system. After eliminating these nonapplicable samples, our final data set consists of 3,057 restaurants. Note that the proportion of nonapplicable restaurants (i.e., restaurants that are permanently closed or have five or fewer photos) in the entire population is 22.45% (hence, a very similar proportion to our sample).[9] The summary statistics of the final data set are available in Table 5.

5.1. Main Results

First, we investigate the performance of the AI-based system in terms of its ability to stimulate user interactions.

Table 5. Summary Statistics

Variable	Observations	Mean	Standard deviation	Min	Max
RestaurantTenure	3,057	189.053	114.660	4	443
NumPhotos	3,057	48.119	143.819	6	4,986
ResRating	3,057	3.906	0.705	1	5
ResPrice	3,057	1.703	0.871	1	5
ResClickTotal	3,057	41.592	152.755	0	4,157
NumPhotoVote	3,057	2.134	18.700	0	553

This analysis allows us to confirm the result from Section 4.3.2 based on the randomized experiment. In this analysis, we follow an approach commonly used in the literature (e.g., Aguiar et al. 2018, Zheng et al. 2019) to apply a logarithm transformation on variables that follow a power law distribution, namely the number of clicks, the number of reviews, and restaurant tenure. Our analyses rely on a panel fixed effect regression specification used by Cavusoglu et al. (2016). Particularly, the regression specification is as follows:

$$\ln(y_{it}+1) = \beta PostImplementation_t + \boldsymbol{\gamma C_{it}} + \rho(t) + \alpha_i + \delta_t + \varepsilon_{it}, \quad (1)$$

where i is a restaurant and t is a specific week. The dependent variable, y_{it}, corresponds to the total number of clicks that restaurant i receives in week t (combined for both the search result page and the restaurant landing page). The term $\boldsymbol{C}$ is a vector of control variables that includes the natural logarithm of the total number of reviews of restaurant i at time t, the average ratings, and the restaurant tenure. The term ρ is the coefficient of linear time trend. The term α_i captures restaurant fixed effects, whereas δ_t captures time fixed effects (at the week level). Lastly, the term ε_{it} represents the error term. The coefficient of interest, β, captures the increase in the number of clicks after the implementation of the AI-based system. In this specification, we use a time window of eight weeks before and eight weeks after the implementation date, so we have a total of $16 \times 3{,}057 = 48{,}912$ observations.

The estimation results are reported in Table 6. We estimate four versions of the model in Equation (1) (with and without time and restaurant fixed effects) and obtain consistent results in all cases. Notably, the estimated coefficient for the variable $PostImplementation_t$ is positive and statistically significant. These results confirm the findings from the field experiment presented in Section 4.3.2. Specifically, the estimation results demonstrate the superior performance of the AI-based system in stimulating user interactions—both statistically and economically.[10]

We conduct additional robustness tests in Online Appendix E.2 as follows. First, we vary the time window (4 and 12 weeks) to check the robustness of the effect using both a shorter-term horizon and a longer-term horizon. Second, we rely on a different identification strategy based on the regression discontinuity in time design, which is commonly used to establish causality for a natural experiment such as the one in our setting. Third, we rely on another identification strategy based on the propensity score matching and difference in differences to alleviate endogeneity concerns and establish causal inferences. Lastly, we perform placebo tests to ensure that the effects we observe are not artifacts of time-correlated confounders (i.e., we did not capture effects that may randomly occur).

Table 6. Impact of the AI-Based System on User Interactions (Metric: Number of Clicks)

Variables	(1) $\ln(y_{it}+1)$	(2) $\ln(y_{it}+1)$	(3) $\ln(y_{it}+1)$	(4) $\ln(y_{it}+1)$
$PostImplementation_t$	0.180*** (0.020)	0.179*** (0.008)	1.450*** (0.400)	1.437*** (0.175)
$\ln(NumberofReviews)_{it}$	1.015*** (0.006)	0.365*** (0.059)	1.015*** (0.006)	0.366*** (0.058)
$AverageRatings_{it}$	0.017** (0.007)	0.056 (0.052)	0.017** (0.007)	0.063 (0.051)
$\ln(RestaurantTenure_{it})$	−0.340*** (0.007)	−0.329*** (0.071)	−0.340*** (0.007)	−0.311*** (0.072)
Linear time trend	−0.003 (0.002)	−0.001 (0.001)	−0.101*** (0.028)	−0.098*** (0.012)
Constant	2.487*** (0.048)	3.349*** (0.402)	1.810*** (0.226)	2.566*** (0.411)
Time fixed effects	No	No	Yes	Yes
Restaurant fixed effects	No	Yes	No	Yes
Observations	48,912	48,912	48,912	48,912
R^2	0.380	0.033	0.386	0.088

Note. Standard errors in parentheses are robust and clustered by restaurant.
**$p < 0.05$; ***$p < 0.01$.

5.2. Performance Heterogeneity

Recall that in the implementation of the AI-based cover image system, there is no other change in the visibility of how the restaurants are shown to the users. In this section, we investigate the heterogeneity in the performance of the AI-based system conditional on the same visibility level, with respect to several restaurant characteristics. For this analysis, we extend the regression specification from Equation (1) by including the interaction terms with the restaurant characteristics of interest. Specifically, the regression specification is as follows:

$$\ln(y_{it}+1) = \beta PostImplementation_t + \boldsymbol{\gamma C_{it}} + \eta(\boldsymbol{X} \times PostImplementation_t) + \rho(t) + \alpha_i + \delta_t + \varepsilon_{it}, \quad (2)$$

where $\boldsymbol{X}$ is a vector of interaction terms that consists of the following characteristics: restaurant tenure (*RestaurantTenure*), number of photos received by each restaurant (*NumPhotos*), average rating of the restaurant (*AverageRatings*), average menu price (*ResPrice*), and the number of clicks that the restaurant attained before the implementation of the AI-based cover image system (*NumClicks*). Similar to the main analysis, we apply a logarithm transformation on *NumPhotos* and *NumClicks* because they follow a power law distribution. The distribution plot of these variables is included in Section D.2 in the online appendix. Note that the restaurant tenure variable is both time and restaurant dependent, so it is also included as an independent term in the specification. Meanwhile, the number of photos received by each restaurant, the average rating of the restaurant, the average menu price, and the number of clicks that the restaurant attained before the implementation of the AI-based cover image system are all calculated prior to the implementation of the AI-based system, so their independent terms are absorbed by the restaurant fixed effects, α_i. The other variables are the same as in Equation (1).

Table 7 presents our estimation results. First, our results show that the main effect (i.e., the positive impact of the AI-based system implementation on user interaction) continues to be positive and statistically significant after accounting for the potential moderating effects of restaurant characteristics. Second, we unveil interesting insights on the heterogeneous impact across different types of restaurants. The interaction effect with $RestaurantTenure_{it}$ is positive and statistically significant, indicating that less recent restaurants tend to benefit more from the implementation of the AI-based system (after controlling for other restaurant characteristics). The coefficient of the interaction term is 0.036, indicating that a restaurant that has been on the platform for one week longer receives 3.6% more clicks after the implementation of the AI-based cover image system. By comparing with the base effect, we can infer that the effect of the AI-based system implementation would double for every approximately 36 weeks that a restaurant stays on the platform. Interestingly, we also find that the implementation of the AI-based system generates more benefits for restaurants with smaller numbers of photos. In other words, our results imply that restaurants that attain a significant crowd engagement level (e.g., number of photos uploaded) before the implementation of the AI-based system receive fewer benefits from the newly implemented system. These two findings bear the following interesting practical implication; our AI-based system allows less recent restaurants that have smaller numbers of user-generated photos to reap higher benefits and hence, potentially close the gap with more established restaurants. In addition, we do not find the heterogeneous effect of the implementation of the AI-based system with respect to the menu price of the restaurants. Furthermore, restaurants with a lower average star rating tend to benefit more from the implementation of the AI-based system. Lastly, restaurants that perform better in attracting clicks before the implementation of the AI-based system attain significantly less benefit from the AI-based system. Knowing the types of restaurants that can benefit from the AI-based system may be useful for platforms that consider adopting this technology. We also provide marginal effects graphs in Section D.3 in the online

Table 7. Heterogeneity of the Impact of the AI-Based System on User Interactions (Metric: Number of Clicks)

	(1) $\ln(y_{it}+1)$	(2) $\ln(y_{it}+1)$	(3) $\ln(y_{it}+1)$	(4) $\ln(y_{it}+1)$	(5) $\ln(y_{it}+1)$
$PostImplementation_t$	1.252***	1.460***	1.424***	1.502***	1.524***
	(0.182)	(0.176)	(0.176)	(0.179)	(0.176)
$\ln(NumberofReviews)_{it}$	0.360***	0.304***	0.364***	0.366***	0.437***
	(0.059)	(0.045)	(0.058)	(0.058)	(0.059)
$AverageRatings_{it}$	0.063	0.057	0.088**	0.053	0.050
	(0.051)	(0.044)	(0.051)	(0.050)	(0.051)
$\ln(RestaurantTenure_{it})$	−0.127	−0.332***	−0.308***	−0.302***	−0.324***
	(0.091)	(0.073)	(0.072)	(0.071)	(0.070)
Linear time trend	−0.100***	−0.098***	−0.098***	−0.098***	−0.098***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
$\ln(RestaurantTenure_{it}) \times PostImplementation_t$	0.036***				
	(0.010)				
$\ln(NumPhotos_{i,t=0}) \times PostImplementation_t$		−0.009***			
		(0.003)			
$ResPrice_i \times PostImplementation_t$			0.007		
			(0.007)		
$ResRatingi, t=0 \times PostImplementation_t$				−0.017**	
				(0.008)	
$\ln(NumClicks_{i,t=0}) \times PostImplementation_t$					−0.036***
					(0.004)
Constant	1.647***	2.480***	2.554***	2.561***	2.563***
	(0.498)	(0.413)	(0.412)	(0.410)	(0.407)
Time fixed effects	Yes	Yes	Yes	Yes	Yes
Restaurant fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	48,912	48,912	48,912	48,912	48,912
R^2	0.089	0.089	0.088	0.088	0.092

Note. Standard errors in parentheses are robust and clustered by restaurant.
**$p < 0.05$; ***$p < 0.01$.

appendix for ease of interpretation, and additional heterogeneity analyses are available in Section E.3 in the online appendix.

5.3. Cover Image Selection

So far, our results have shown that the AI-based system can select cover images that generate higher user interactions relative to the crowd-based system. One may naturally wonder what the underlying mechanism is that leads to such a superior performance. In this section, we analyze how the cover images are selected by the crowd and by AI.

5.3.1. AI Vs. Crowd in Selecting Cover Images. We begin by drawing upon a conceptual model proposed by Pomerol (1997), which suggests that AI tends to rely more on direct information (i.e., image characteristics such as color and composition), whereas humans tend to utilize both direct information and peripheral factors (i.e., nonimage-related characteristics such as those of the image uploader's) when processing tasks. In line with this conceptual framework, we explore whether the use of peripheral factors by the crowd augments or deteriorates the use of direct information compared with how AI uses the information. For this analysis, we use a sample of user-generated photos from the 3,057 restaurants in our data set. These photos are organized at the restaurant level, where we identify two sets of cover images. The first set comprises the cover images of each restaurant on the day that the AI-based system was implemented (i.e., postimplementation), whereas the second set is those cover images of each restaurant one day before the implementation day (i.e., preimplementation). Each set of cover images has $3{,}057 \times 5 = 15{,}285$ images. Among these 30,570 images, 5,825 of them were selected as cover images by both systems. Therefore, our image data set consists of $[(15{,}285 - 5{,}825) \times 2] + 5{,}825 = 24{,}745$ unique images. For all the photos that were selected as cover images, we derive several variables that represent direct information and peripheral factors.

First, we construct variables that represent image characteristics. On the basis of theoretical foundations grounded in the photography literature (e.g., Datta et al. 2006), pioneer studies in management research have empirically established several quantifiable image characteristics that influence human perception (Zhang et al. 2022). We employ machine learning-based feature extraction techniques to retrieve several photographic attributes for each image that capture the intrinsic image characteristics. We consider the following two representative categories related to intrinsic image characteristics.

The first category is related to the color components of the image. Prior research has found color to have a significant influence on image aesthetic and attractiveness, especially in the context of food images (e.g., Nishiyama et al. 2011). We consider the following attributes.

- *Brightness* captures the average of the value dimension in hue, saturation, and value (HSV) across pixels (Datta et al. 2006). A higher brightness refers to a brighter image.
- *Saturation* is the average of the saturation dimension of HSV across pixels, and it refers to the color purity of the image (Datta et al. 2006). A higher saturation score refers to a more saturated image, which is known to enhance an image's attractiveness (Vernon and Bartel 1985).

To calculate the values, we first convert each image from the red, green, and blue space to HSV and extract each attribute accordingly. All attributes are normalized so that values lie between zero and one.

The second category is related to the composition components of the image. The composition of each image generally captures the arrangement of visual subjects and visual elements within the image. Previous studies have shown that images with good composition tend to reduce viewers' cognitive demands and allow them to focus more readily (Grill and Scanlon 1990). In this context, we consider the following attributes.

- *Diagonal dominance* measures how close the salient region of the photo is positioned relative to the diagonal lines. The distances between the center of the main object with each of the two diagonals are calculated. We then define the diagonal dominance as the negative of the smallest value of both distances (Wang et al. 2013).
- The *rule of thirds* is a principle for composing good visual images. If the main element of the photo is positioned near the intersection of two vertical lines and two horizontal lines of the image, the photo has good composition. The distances between the center of the main object (e.g., the most salient region) and the intersections of two vertical and two horizontal lines are calculated. Then, we define the rule of thirds as the negative of the minimum of the four distances. A higher score means a stronger rule of thirds.

To calculate these values, we measure a saliency score of each pixel in the image by using a computer vision algorithm (Montabone and Soto 2010). Based on the saliency score, we divide the image into 10 segments based on the superpixel algorithm (Ren and Malik 2003). Then, we identify the segment with the highest average saliency as the salient region. Finally, all attributes are normalized such that the values are between zero and one.

Second, we construct variables that represent peripheral factors: that is, characteristics not directly related to the images that may influence user perception. We identify the following three types of variables that belong to this category:

The first type of peripheral factors is related to the user who posted the image. Prior literature has shown that the characteristics of the individual who generates the content, particularly the popularity of the user, can significantly influence the interaction level of the content (Yu et al. 2023). In the context of user-generated photos, such a popularity effect may also play a significant role in the crowd decision process when selecting and interacting with images. Thus, we consider the average number of followers of the user who posted the cover image and the average number of comments received by this user as proxies for the popularity effect.

The second type of peripheral factors is the exposure level of the image. Psychology literature has shown that two copies of the same content may be perceived differently depending on their exposure level. This mere exposure effect (Zajonc 2001) is widely adopted to explain multiple phenomena that could affect users' perception and interactions, including the familiarity effect (i.e., more familiar content tends to be viewed more positively), the bandwagon effect (i.e., perceived quality of user-generated content becomes higher when it receives a higher level of public attention), and the conformity effect (i.e., platform users tend to herd toward the direction set by public sentiments). In our empirical context, such a mere exposure effect may also influence how images are selected and interacted with. Thus, we calculate the tenure of the photos as a proxy to capture the mere exposure effect.

The third type of peripheral factors relates to the consistency (or inconsistency) of the images. When individuals make decisions, they tend to rely heavily on the first piece of information offered (i.e., "anchoring") (Fresneda and Gefen 2019). Cover images are representative images of the restaurant that are exposed to the users before they evaluate the images on the platform. Accordingly, the variation in cover images, which is an initial set of information, could affect the subsequent perception of users as well as their interactions. Although it is widely known that consumers tend to have heterogeneous preferences (Zhang and Sarvary 2015), there exist theoretical tensions in the literature on whether consumers prefer consistent or inconsistent content (Zheng et al. 2018). We empirically explore this issue by incorporating object similarity among cover images (i.e., the similarity of the objects that appear in each image and in each set of cover images) and the average standard deviation of intrinsic characteristics among cover images. Here, object similarity measures the object-level (ingredient-level) similarity between feature images. To identify the contents of the photos, we leveraged the Clarifai API, which was among the top five winners in the ImageNet 2013 competition. In particular, Clarifai provides a food classification model that recognizes more than 1,000 food items. Using this API, we extracted 10 labels with the prediction score from each image. To consider semantic

similarity across five feature images, we first calculate the cosine similarity of two photos of the feature image set considering all of the image combinations. Then, we calculate the average value of the (5–1)! cosine similarity values. Following that, we rely on a paired *t*-test to examine the difference between these factors preimplementation (i.e., under the crowd-based system) versus postimplementation (i.e., under the AI-based system).

Table 8 reports the estimated results. Interestingly, the intrinsic characteristics of the cover images selected by the AI-based system are significantly higher relative to the cover images selected by the crowd-based system. In addition, our results reveal that the use of peripheral factors is also significantly different between the two systems when selecting cover images. Particularly, the crowd tends to select images that are uploaded by popular users and are less recent. In the same vein, the crowd is less consistent in the selection of cover images both in terms of object similarity in the images and in intrinsic image characteristics variation. In other words, cover images selected by AI have better intrinsic characteristics, whereas the images selected by the crowd are more influenced by peripheral factors, such as the popularity of image uploaders and the tenure of the photos. Furthermore, the sets of cover images selected by the AI-based system are more consistent. Our findings indicate that the use of peripheral factors by humans in selecting cover images does not augment their use of intrinsic characteristics. As a result, cover images selected by AI are of significantly better intrinsic characteristics.

5.3.2. Cover Images and User Interactions. We next delve into the role of intrinsic image characteristics and peripheral factors on user interactions. For this analysis, we perform a regression based on the following specification:

$$\ln(y_{i,\,PostImplementation}) - \ln(y_{i,\,PreImplementation}) = \boldsymbol{\beta}(\boldsymbol{X}_{i,\,PostImplementation} - \boldsymbol{X}_{i,\,PreImplementation}) + (u_{i,\,PostImplementation} - u_{i,\,PreImplementation}), \quad (3)$$

where y_i is the number of clicks on the cover images of restaurant i and $\boldsymbol{X}_i$ is the vector of intrinsic image characteristics and peripheral factors of cover images for restaurant i. Essentially, the regression in Equation (3) analyzes whether the change in the number of clicks after the implementation of the AI-based system can be explained by the differences in intrinsic image characteristics and peripheral factors of the cover images. We note that some of these variables are naturally correlated with others. We thus remove the variables that are strongly correlated with other variables to alleviate multicollinearity issues. Specifically, we remove the number of comments received (high correlation with the number of followers) and the average standard deviation of image characteristics (high correlation with object similarity). The regression results are presented in Table 9.

We find that the estimated coefficients related to intrinsic characteristics are all positive and statistically significant, thus indicating that users interact more with images that have better intrinsic characteristics. Such positive effects are observed for both color and composition components, which are in line with prior photographic studies (Datta et al. 2006). After controlling for intrinsic image characteristics, our findings reveal that users tend to interact less with photos uploaded by popular users (i.e., negative coefficient of ln(*NumFollower*)). In addition, the tenure of the photo does not seem to have a significant impact on user interactions. The discrepancy in the factors that influence image selection and the factors that influence user interactions is particularly insightful. Finally, our results show that users tend to react positively when cover images of the same restaurant are more consistent (i.e., positive coefficient of *ObjectSimilarity*). Such a stark contrast can also explain the difference in performance of the AI-based system relative to the crowd-based system in stimulating user interactions.

5.3.3. How Does the Crowd Select Cover Images? Our results from Section 5.3.1 demonstrate that the crowd tends to rely more on peripheral factors and less on intrinsic image characteristics than AI. In this subsection,

Table 8. Photo-Level Analysis

	Preimplementation mean	Postimplementation mean	Mean differences	Percentage differences	*p*-value
Intrinsic image characteristics					
Brightness	0.572	0.600	0.028	4.985	<0.001
Saturation	0.385	0.395	0.010	2.679	<0.001
Rule of thirds	0.594	0.603	0.009	1.530	<0.001
Diagonal distance	0.706	0.718	0.012	1.699	<0.001
Peripheral factors					
Number of followers	3,764	2,424	−1,340	−35.606	<0.001
Number of comments received	0.665	0.397	−0.268	−40.333	<0.001
Photo tenure	965.044	890.018	−75.025	−7.774	<0.001
Object similarity	0.355	0.397	0.042	11.827	<0.001
Average standard of image characteristics	0.495	0.474	−0.021	−4.214	<0.001

we provide an additional analysis that specifically examines how the crowd selects cover images under the crowd-based cover image system. Intrinsic image characteristics and peripheral factors that we have considered in the previous section may affect the crowd voting decision, which broadly consists of three steps within the platform. First, users are exposed to restaurant-level information when they first access the website. Their voting decisions may rely on the first piece of information they have been exposed to (i.e., third peripheral factors). Given that *ObjectSimilarity* and *Avg.Std.of image Characteristics* cannot be captured at the image level, we have included an alternative proxy variable *Feature* (i.e., whether the image is presently a featured image). Furthermore, we control for several restaurant characteristics, such as the number of reviews and the average rating, that may naturally affect voting decisions. Then, after accessing the restaurant page, individuals who explore the photos will first see the most recent picture. As the exposure level of the image may influence individuals' decisions (i.e., second peripheral factors), we have included the variable *PhotoTenure*. Lastly, the crowd may consider the information on the photo uploaders (e.g., their popularity—first peripheral factors) as well as intrinsic image characteristics when making voting decisions.

For this analysis, we use a sample of 132,865 user-generated photos that were uploaded to the platform before the implementation of the AI-based system from the 3,057 restaurants in our data set. Essentially, our analysis is an image-level analysis, where the outcome variable is the number of votes that each image receives. Similar to our main analyses, the outcome variable is log transformed. Table 10 reports the estimation results of the voting decision by the crowd. The estimation results from an alternative specification based on a negative binomial model are available in Section E.2.5 in the online appendix.

We find that platform users use only two intrinsic image characteristics, *brightness* and *saturation*, in their cover image selection. Meanwhile, most of the peripheral factors are statistically significant. Namely, users tend to vote for images uploaded by popular users, images of popular restaurants, images that are already cover images, and recent images. Notably, these peripheral factors are commonly used by review platforms, and it is practically difficult to present images to the users without these factors or to prevent platform users from using them in the selection process. These findings, in combination with our results from Section 5.3.1, showcase the mechanisms behind the superior performance of the AI-based system in its ability to select cover images that boost user interactions.

Table 9. User Engagement, Intrinsic Image Characteristics, and Peripheral Factors

	Difference in number of clicks
Brightness	0.614*** (0.108)
Saturation	0.343*** (0.109)
Rule of Thirds	0.293*** (0.106)
Diagonal Distance	0.137* (0.075)
ln(*NumFollower*)	−0.020*** (0.004)
ln(*PhotoTenure*)	−0.012 (0.016)
Object Similarity	0.228*** (0.053)
Constant	3.969*** (0.149)
Observations	6,114
R^2	0.052

Note. Standard errors are in parentheses.
*$p < 0.1$; ***$p < 0.01$.

Table 10. The Selection of Cover Images by the Crowd

	ln(*NumPhotoVotes*+1)
Brightness	−0.020*** (0.003)
Saturation	0.013*** (0.002)
Rule of Thirds	−0.003 (0.002)
Diagonal Distance	0.002 (0.002)
ln(*NumFollowers*)	0.003*** (0.001)
ln(*PhotoTenure*)	−0.025*** (0.001)
Feature	0.004*** (0.001)
ln(*NumReviews*)	0.002*** (0.001)
AvgRatings	−0.004*** (0.001)
Constant	0.129*** (0.004)
Observations	132,865
R^2	0.053

Note. Standard errors are in parentheses.
***$p < 0.01$.

In summary, our results demonstrate the positive impact of adopting an AI-based cover image system on the number of clicks generated by users. This measure is considered as the primary variable of interest by the platform because it has a direct influence on the platform's advertising revenue, which is the main revenue stream of the platform. At the same time, one may raise the concern that the increase in the number of clicks may come at a cost of other important user participation metrics that may negatively affect the platform in the long run. For example, the adoption of the AI-based cover image system may discourage users to upload photos or vote on them because such activities are less relevant now that AI is at the wheel. We investigate such a potential cannibalization effect in Section E.4 in the online appendix. In short, our results demonstrate that such a concern is not statistically significant.

6. Conclusions and Implications

With the increasing ubiquity of the mobile economy, user-generated photos have become an important component of online review platforms. These photos clearly provide additional information that is not available in textual contents. This is especially true for review platforms related to hotels and restaurants. In this context, an important challenge faced by most platforms is how to organize and display these images. Unlike the textual content part of the reviews, which can be handled with

ease thanks to recent advances in text analytics, images are significantly more difficult to manage because administrator interventions are frequently required. This issue is particularly important for review platforms given that these user-generated photos are directly linked to user satisfaction and to advertising revenue.

A common approach to handling user-generated photos is to rely on the users themselves. Although such a crowd-based system may perform well, the reliance on users and administrators can affect the platform's sustainability in the long run. In this paper, we collaborate with a large restaurant review platform in Asia to develop an alternative approach based on a deep learning model to evaluate user-generated photos. We first ran a randomized field experiment and show that the AI-based system outperforms the traditional crowd-based system in terms of user interactions (measured by clicks). Following the success of the field experiment, the platform decided to fully deploy the AI-based system. We then leverage the resulting observational data to conduct various empirical analyses. We find that the benefit of the AI-based system over the crowd-based system is higher for restaurants with a longer tenure, a limited number of user-generated photos, a lower star rating, and lower user engagement during the crowd-based era. We also show that the difference in the performance can be explained by the factors used to select images; AI tends to rely on intrinsic image characteristics, whereas the crowd tends to use both intrinsic characteristics and peripheral factors, such as the popularity of users who upload the images. Lastly, we verify that the increase in the number of clicks after adopting the AI-based system does not come at a cost of decreasing other important user participation metrics, such as photo votes and photo uploads. The findings that the implementation of the AI-based cover image system leads to a significant increase in the number of clicks, which is a key operational metric that has direct implications on platform revenues, while retaining the same level of other user engagements, such as photo votes and photo uploads, reinforce the platform's confidence and satisfaction on the AI-based system. As such, the partner platform views the AI-based system as an integral part of the platform's future operations and continues to utilize it in the production environment more than three years after the implementation.

Our findings bear several research implications. Prior studies have shown that a voluntarily engaged crowd is effective in quality evaluation. In this paper, we empirically demonstrate that a deep learning image selection model significantly outperforms the crowd in selecting images. The fact that we compare AI versus the crowd (as opposed to a single agent) distinguishes our study from prior work on user-generated content. Our subsequent analyses also reveal the reason that AI outperforms the crowd. In that regard, our work paves the way for an emerging research stream at the interface of AI and the crowd.

This research also provides strategic insights on how platforms can effectively monetize user-generated content in the context of cover images. The positive effects of the AI-based system on user interactions offer an opportunity to successfully monetize visual content for advertising. Especially, platforms that use cost-per-click advertising models can significantly benefit from the increase in the number of clicks observed in our results. In such a case, the increase in the number of clicks directly translates into a revenue increase. Notably, our partner platform continues to use the AI-based system as of today, which also testifies to the benefits of this system in the long run. Moreover, our results illustrate that the AI-based system performs better at identifying intrinsic characteristics, whereas the crowd tends to be more influenced by peripheral factors. Accordingly, AI outperforms the crowd in platforms where users are influenced by intrinsic characteristics but not by peripheral factors, such as platforms that cater to users' information searches. Meanwhile, it is plausible that the crowd may outperform AI if the peripheral factors can augment the use of intrinsic characteristics or have a positive influence on user interactions. Platforms with such characteristics, such as social influence-based platforms, may benefit from a crowd-based system or from a hybrid system that combines AI and the crowd.

Finally, our deep learning-based system serves as a useful reference for a practical system that can quantitatively assess unstructured content with business-oriented outcomes, which has received a growing interest in recent years. Our model is readily reproducible because it is primarily trained on publicly available data. Further, the transfer learning process used in this paper requires minimal human intervention. As such, the AI-based model presented in this paper presents a viable alternative even for smaller platforms that are interested in implementing an AI solution under limited resources.

Acknowledgments

The authors thank the department editor, the associate editor, and the anonymous referees for their insightful comments, which have helped improve this paper. The authors also thank the partner company for the collaboration opportunity and for sharing the data.

Endnotes

[1] In our study, MobileNets completes the classification task at least 10% faster compared with other classifiers, whereas the decrease in the classification accuracy is less than one percentage point.

[2] As we will describe at the end of this section, the images in our data set are rated by skilled human agents with a score between 1 and 10. Hence, the task of the AI model is to output a multilabel classification where the potential labels are 1, 2, …, 10. As a result, our AI model needs to have 10 output nodes to accommodate such a requirement.

[3] We conduct a laboratory experiment to let student subjects rate the images and find that ratings issued by skilled human agents are highly correlated with those issued by undergraduate students.

Thus, it is unlikely that the results we observe are solely driven by the labels provided by skilled human agents.

[4] The randomization process relies on a round-robin procedure with a filter where users who were already redirected to the experimental platform once would not be redirected again to prevent duplication. This filtering rule applies to both users who log in and user who do not log in. There are two methods that the platform uses to track users' browsing history for users who do not log in. First, for users who use the platform's mobile application, the platform collects the device identification number and uses it to identify unauthenticated users. Second, for users who use an internet browser, the platform implicitly profiles unauthenticated users based on their IP address, session identification, cookies, screen size, etc. Although this profiling is not perfect, it is reasonably accurate to uniquely identify unauthenticated users.

[5] A session identification is a 128-bit hexadecimal string randomly generated by a web server. The platform performed the *div* operation on the last bit of the session identification with the value *a*. As such, the result (i.e., remaining quotient) is either zero (for session identifications ending with zero to nine) or one (for session identifications ending with a–f). The identifications with the remaining quotient of zero (one) observed the crowd-based (AI-based) cover image system.

[6] Based on historical data, the platform considers the following actions as desirable user-restaurant interactions: (1) save the restaurant to their personal bookmark, (2) view the map of the restaurant, (3) get the directions to the restaurant address, (4) call the restaurant, and (5) share the restaurant on their social network. During the experiment, the platform only collected the data regarding these actions in an aggregate fashion (i.e., the platform did not collect data regarding each action separately because of data privacy regulations as discussed earlier).

[7] We also use observational data to plot the distribution of the increase in click ratio that we observe in the randomized field experiment. The plots are available in Section D.1 in the online appendix.

[8] The platform retains reviews of restaurants that are permanently closed, as is the case with other review platforms such as Yelp.

[9] Among the 3,057 restaurants in our final data set, 574 of them use a tiebreaking mechanism to select cover images in situations where multiple photos have the same number of user votes. During our study period and before the implementation of the AI-based system, the tiebreaking mechanism used by the platform relied on an aesthetic score of each photo generated by a third-party application programming interface (API). To ensure a clean identification, we exclude these 574 restaurants and rerun the analysis. The results, reported in Section E.2.1 in the online appendix, are qualitatively similar to our main results.

[10] Note that the purpose of our econometrics model is to identify the relationship between $PostImplementation_t$ and the number of clicks and that it is not to explain how the number of clicks is generated. Hence, even when the R^2 value of a model is low (i.e., a model has low explanatory power on the variance of the outcome variable), the relationship between the independent variable and the dependent variable that the model establishes with statistical significance continues to be valid. In fact, the R^2 of our models is comparable with that of other empirical studies that utilize fixed effects models (e.g., Cao et al. 2022).

References

Adamopoulos P, Ghose A, Todri V (2018) The impact of user personality traits on word of mouth: Text-mining social media platforms. *Inform. Systems Res.* 29(3):612–640.

Adulyasak Y, Benomar O, Chaouachi A, Cohen MC, Khern-am-nuai W (2023) Using AI to detect panic buying and improve products distribution amid pandemic. *AI Soc.*, ePub ahead of print April 15, https://doi.org/10.1007/s00146-023-01654-9.

Aguiar L, Claussen J, Peukert C (2018) Catch me if you can: Effectiveness and consequences of online copyright enforcement. *Inform. Systems Res.* 29(3):656–678.

Aouad A, Saban D (2023) Online assortment optimization for two-sided matching platforms. *Management Sci.* 69(4):2069–2087.

Bai B, Dai H, Zhang DJ, Zhang F, Hu H (2022) The impacts of algorithmic work assignment on fairness perceptions and productivity: Evidence from field experiments. *Manufacturing Service Oper. Management* 24(6):3060–3078.

Cao X, Zhang D, Huang L (2022) The impact of the Covid-19 pandemic on the behavior of online gig workers. *Manufacturing Service Oper. Management* 24(5):2611–2628.

Cavusoglu H, Phan TQ, Cavusoglu H, Airoldi EM (2016) Assessing the impact of granular privacy controls on content sharing and disclosure on Facebook. *Inform. Systems Res.* 27(4):848–879.

Cheng YH, Ho HY (2015) Social influence's impact on reader perceptions of online reviews. *J. Bus. Res.* 68(4):883–887.

Cohen MC (2018) Big data and service operations. *Production Oper. Management* 27(9):1709–1723.

Cohen MC, Fiszer MD, Kim BJ (2022) Frustration-based promotions: Field experiments in ride-sharing. *Management Sci.* 68(4):2432–2464.

Cohen MC, Fiszer MD, Ratzon A, Sasson R (2023) Incentivizing commuters to carpool: A large field experiment with Waze. *Manufacturing Service Oper. Management* 25(4):1263–1284.

Cui R, Li M, Zhang S (2022) AI and procurement. *Manufacturing Service Oper. Management* 24(2):691–706.

Cui Y (2020) *Artificial Intelligence and Judicial Modernization* (Springer, Berlin).

Datta R, Joshi D, Li J, Wang JZ (2006) Studying aesthetics in photographic images using a computational approach. Leonardis A, Bischof H, Pinz A, eds. *Eur. Conf. Comput. Vision* (Springer, Berlin), 288–301.

Feldman P, Frazelle AE, Swinney R (2023) Managing relationships between restaurants and food delivery platforms: Conflict, contracts, and coordination. *Management Sci.* 69(2):812–823.

Fresneda JE, Gefen D (2019) A semantic measure of online review helpfulness and the importance of message entropy. *Decision Support Systems* 125:113117.

Fügener A, Grahl J, Gupta A, Ketter W (2022) Cognitive challenges in human–artificial intelligence collaboration: Investigating the path toward productive delegation. *Inform. Systems Res.* 33(2):678–696.

Gallino S, Moreno A (2018) The value of fit information in online retail: Evidence from a randomized field experiment. *Manufacturing Service Oper. Management* 20(4):767–787.

Ghadiyaram D, Bovik AC (2015) Massive online crowdsourced study of subjective and objective picture quality. *IEEE Trans. Image Processing* 25(1):372–387.

Grill T, Scanlon M (1990) *Photographic Composition* (Amphoto Books, New York).

Guo J, Zhang W, Fan W, Li W (2018) Combining geographical and social influences with deep learning for personalized point-of-interest recommendation. *J. Management Inform. Systems* 35(4):1121–1153.

Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H (2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications. Preprint, submitted April 17, https://arxiv.org/abs/1704.04861.

Huang N, Sun T, Chen P, Golden JM (2019) Word-of-mouth system implementation and customer conversion: A randomized field experiment. *Inform. Systems Res.* 30(3):805–818.

Iandola FN, Moskewicz MW, Ashraf K, Keutzer K (2016) Firecaffe: Near-linear acceleration of deep neural network training on compute clusters. Bajcsy R, Li F-F, Tuytelaars T, eds. *Proc. IEEE Conf. Comput. Vision Pattern Recognition* (IEEE, Piscataway, NJ), 2592–2600.

Järvelin K, Kekäläinen J (2000) IR evaluation methods for retrieving highly relevant documents. Harman D, Kelly D, eds. *ACM*

SIGIR Conf. Res. Development Inform. Retrieval (Association for Computing Machinery, New York), 41–48.

Keding C (2021) Understanding the interplay of artificial intelligence and strategic management: Four decades of research in review. *Management Rev. Quart.* 71:91–134.

Khern-am-nuai W, Ghasemkhani H, Qiao D, Kannan K (2023a) The impact of online Q&As on product sales: The case of Amazon answer. *Inform. Systems Res.*, ePub ahead of print June 6, https://doi.org/10.1287/isre.2023.1233.

Khern-am-nuai W, Hashim MJ, Pinsonneault A, Yang W, Li N (2023b) Augmenting password strength meter design using the elaboration likelihood model: Evidence from randomized experiments. *Inform. Systems Res.* 34(1):157–177.

Kittur A, Chi E, Pendleton BA, Suh B, Mytkowicz T (2007) Power of the few vs. wisdom of the crowd: Wikipedia and the rise of the bourgeoisie. Holmquist LE, Brown B, eds. Alt.CHI (Association for Computing Machinery, New York), 1–9.

Koh TK (2019) Adopting seekers' solution exemplars in crowdsourcing ideation contests: Antecedents and consequences. *Inform. Systems Res.* 30(2):486–506.

Kohavi R, Thomke S (2017) The surprising power of online experiments. *Harvard Bus. Rev.* 95(5):74–82.

Kokkodis M, Lappas T (2020) Your hometown matters: Popularity-difference bias in online reputation platforms. *Inform. Systems Res.* 31(2):412–430.

Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. Pereira F, Burges CJ, Bottou L, Weinberger KQ, eds. *Adv. Neural Inform. Processing Systems* (Association for Computing Machinery, New York), 1097–1105.

Kumar S, Mookerjee V, Shubham A (2018) Research in operations management and information systems interface. *Production Oper. Management* 27(11):1893–1905.

Kyung N, Kwon HE (2022) Rationally trust, but emotionally? The roles of cognitive and affective trust in laypeople's acceptance of ai for preventive care operations. *Production Oper. Management*, ePub ahead of print June 30, https://doi.org/10.1111/poms.13785.

Lee D, Hosanagar K (2019) How do recommender systems affect sales diversity? A cross-category investigation via randomized field experiment. *Inform. Systems Res.* 30(1):239–259.

Liu QB, Karahanna E (2017) The dark side of reviews: The swaying effects of online product reviews on attribute preference construction. *MIS Quart.* 41(2):427–448.

Lou J, Yang H (2018) Food image aesthetic quality measurement by distribution prediction. Working paper, Stanford University, Stanford, CA.

Manshadi V, Rodilitz S, Saban D, Suresh A (2022) Online algorithms for matching platforms with multi-channel traffic. Preprint, submitted April 22, http://dx.doi.org/10.2139/ssrn.4036904.

Miller A (2016) Finding beautiful yelp photos using deep learning. Accessed November 3, 2021, https://engineeringblog.yelp.com/2016/11/finding-beautiful-yelp-photos-using-deep-learning.html.

Mithas S, Chen ZL, Saldanha TJ, De Oliveira Silveira A (2022) How will artificial intelligence and industry 4.0 emerging technologies transform operations management? *Production Oper. Management* 31(12):4475–4487.

Montabone S, Soto A (2010) Human detection using a mobile platform and novel features derived from a visual saliency mechanism. *Image Vision Comput.* 28(3):391–402.

Nishiyama M, Okabe T, Sato I, Sato Y (2011) Aesthetic quality classification of photographs based on color harmony. Boult T, Kanade T, Peleg S, eds. *IEEE Conf. Comput. Vision Pattern Recognition* (IEEE Computer Society, Piscataway, NJ), 33–40.

Otterbacher J (2009) 'Helpfulness' in online communities: A measure of message quality. Olsen DR, Arthur RB, eds. *Proc. SIGCHI Conf. Human Factors Comput. Systems* (Association for Computing Machinery, New York), 955–964.

Overgoor G, Rand W, Dolen WV (2020) The champion of images: Understanding the role of images in the decision-making process of online hotel bookings. Bui T, ed. *Proc. 53rd Hawaii Internat. Conf. System Sci.* (University of Hawai'i at Mānoa, Honolulu, HI), 4069–4078.

Pan SJ, Yang Q (2009) A survey on transfer learning. *IEEE Trans. Knowledge Data Engrg.* 22(10):1345–1359.

Park K, Hong S, Baek M, Han B (2017) Personalized image aesthetic quality assessment by joint regression and ranking. Medioni G, Michael D, Sarkar S, eds. *2017 IEEE Winter Conf. Appl. Comput. Vision (WACV)* (IEEE Computer Society, Piscataway, NJ), 1206–1214.

Pomerol JC (1997) Artificial intelligence and human decision making. *Eur. J. Oper. Res.* 99(1):3–25.

Ponomarenko N, Ieremeiev O, Lukin V, Egiazarian K, Jin L, Astola J, Vozel B, et al. (2013) Color image database TID2013: Peculiarities and preliminary results. Beghdadi A, ed. *Eur. Workshop Visual Inform. Processing (EUVIP)* (IEEE, Piscataway, NJ), 106–111.

Ren X, Malik J (2003) Learning a classification model for segmentation. Triggs B, Zisserman A, ed. *IEEE Internat. Conf. Comput. Vision* (IEEE, Piscataway, NJ), 10–17.

Rubner Y, Tomasi C, Guibas LJ (1998) A metric for distributions with applications to image databases. Chandran S, Desai U, eds. *Sixth Internat. Conf. Comput. Vision (IEEE Catalog No. 98CH36271)* (IEEE, Piscataway, NJ), 59–66.

Shin D, He S, Lee GM, Whinston AB, Cetintas S, Lee KC (2020) Enhancing social media analysis with visual data analytics: A deep learning approach. *MIS Quart.* 44(4):1459–1492.

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. Preprint, September 4, https://arxiv.org/abs/1409.1556.

Spring M, Faulconbridge J, Sarwar A (2022) How information technology automates and augments processes: Insights from artificial-intelligence-based systems in professional service operations. *J. Oper. Management* 68(6–7):592–618.

Sun T, Viswanathan S, Zheleva E (2021) Creating social contagion through firm-mediated message design: Evidence from a randomized field experiment. *Management Sci.* 67(2):808–827.

Surowiecki J (2004) *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business* (Doubleday, New York).

Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z (2016) Rethinking the inception architecture for computer vision. Bajcsy R, Li F-F, Tuytelaars T, eds. *Proc. IEEE Conf. Comput. Vision Pattern Recognition* (IEEE Computer Society, Piscataway, NJ), 2818–2826.

Talebi H, Milanfar P (2018) Nima: Neural image assessment. *IEEE Trans. Image Processing* 27(8):3998–4011.

Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C (2018) A survey on deep transfer learning. Kůrková V, Manolopoulos Y, Hammer B, Iliadis L, Maglogiannis I, eds. *Internat. Conf. Artificial Neural Networks* (Springer, Berlin), 270–279.

Tan TF, Netessine S (2020) At your service on the table: Impact of tabletop technology on restaurant performance. *Management Sci.* 66(10):4496–4515.

Vernon R, Bartel D (1985) Effect of hue, saturation, and intensity on color selection by the onion fly, *Delia antiqua* (meigen)(diptera: Anthomyiidae) in the field. *Environ. Entomology* 14(3):210–216.

Wang Y, Goes P, Wei Z, Zeng D (2019) Production of online word-of-mouth: Peer effects and the moderation of user characteristics. *Production Oper. Management* 28(7):1621–1640.

Wang Y, Wang L, Li Y, He D, Liu TY (2013) A theoretical analysis of NDCG ranking measures. Shalev-Shwartz S, Steinwart I, eds.

Proc. 26th Annual Conf. Learn. Theory (COLT) (Proceedings of Machine Learning Research (PMLR), New York), 25–54.

Xu Y, Armony M, Ghose A (2021) The interplay between online reviews and physician demand: An empirical investigation. *Management Sci.* 67(12):7344–7361.

Xu Y, Lu B, Ghose A, Dai H, Zhou W (2023) The interplay of earnings, ratings, and penalties on sharing platforms: An empirical investigation. *Management Sci.*, ePub ahead of print April 19, https://doi.org/10.1287/mnsc.2023.4761.

Yu Y, Khern-am-nuai W, Pinsonneault A, Wei Z (2023) The impacts of social interactions and peer evaluations on online review platforms. *J. Management Inform. Systems* Forthcoming.

Zajonc RB (2001) Mere exposure: A gateway to the subliminal. *Current Directions Psych. Sci.* 10(6):224–228.

Zhang K, Sarvary M (2015) Differentiation with user-generated content. *Management Sci.* 61(4):898–914.

Zhang Q, Yang LT, Chen Z, Li P (2018) A survey on deep learning for big data. *Inform. Fusion* 42:146–157.

Zhang S, Lee D, Singh PV, Srinivasan K (2022) What makes a good image? Airbnb demand analytics leveraging interpretable image features. *Management Sci.* 68(8):5644–5666.

Zheng J, Qi Z, Dou Y, Tan Y (2019) How mega is the mega? Exploring the spillover effects of WeChat using graphical model. *Inform. Systems Res.* 30(4):1343–1362.

Zheng X, Hong Y, Ren X, Cao J, Yang S (2018) Information inconsistencies in multi-dimensional rating systems. Baskerville R, Nickerson R, eds. *39th Internat. Conf. Inform. Systems* (Association for Information Systems, Atlanta), 1–17.