

## Human Decision Making in AI Augmented Systems: Evidence from the Initial Coin Offering Market

Saunak Basu  
University of Illinois Urbana-  
Champaign  
[saunakb2@illinois.edu](mailto:saunakb2@illinois.edu)

Aravinda Garimella  
University of Illinois Urbana-  
Champaign  
[aravinda@illinois.edu](mailto:aravinda@illinois.edu)

Wencui Han  
University of Illinois Urbana-  
Champaign  
[wenhan@illinois.edu](mailto:wenhan@illinois.edu)

Alan R. Dennis  
Indiana University  
Bloomington  
[ardennis@indiana.edu](mailto:ardennis@indiana.edu)

### Abstract

*The growing consensus that human intelligence and artificial intelligence are complementary has led to Human-AI hybrid systems. As digital platforms incorporating human-AI hybrids, platform designers need to evaluate the influence of AI on human judgment, and how such hybrid systems perform. In this paper, we investigate: Are human decisions influenced by AI agents in high uncertainty environments, such as evaluating ICO projects? Under what situations are humans able to mitigate AI agents-induced errors? Our results suggest that in general, humans are influenced by AI agents. Humans tend to use AI as a filter to rule out low quality projects, while a high AI rating triggers human expert to apply their own judgment.*

### 1. Introduction

With the advancing of Artificial Intelligence (AI) technology, human AI hybrid systems are increasingly been used in various industries, such as health care, finance, e-commerce, politics and government, and much more (Fethi and Pasiouras 2010 [1], Esteva et al. 2017 [2], Leachman and Merlino 2017) [3]. However, research has yet to understand how these systems perform, and how humans may be influenced by AI agents. The spectrum of the involvement of AI in the completion of a task can range from task augmentation, task assemblage, to task substitution (Constantinides 2019) [4]. In this paper, we are interested in understanding how do AI agents influence human decision making in an AI augmented system?

We focus on a context where the uncertainty is high, and the decision making is complex – the evaluation of Initial Coin Offerings (ICO). ICO is a fundraising mechanism by which startup companies sell crypto tokens in exchange for traditional fiat currency. We collected data from a leading ICO evaluation platform, where an AI agent first evaluates and rates ICOs based on an algorithm that takes into

consideration observable and quantitative properties of the ICO. After this, human experts evaluate the ICO and provide their ratings.

The return from investing in the ICO tokens depends on the eventual success of the projects, which depends on a myriad of factors; some of them are readily observable, while many other aspects such as vision and potential are not easily quantifiable. While AI algorithm can effectively collect information and perform consistently without the influence of emotions or biases, humans possess tacit knowledge that cannot be explicitly explained, especially in evaluating aspects like vision or potential. However, there is little empirical evidence on how human experts incorporate AI agent assessments into their decision making. In this paper, we are interested in understanding the following research questions. Is human decision making influenced by AI agents in a human-AI hybrid system in high-risk environments with high-failure probability? Does this depend upon whether the AI agent predicts success or failure?

Our analysis finds that overall, human experts are influenced by the AI agent. When the AI agent rates a project low, human expert are likely to align with the agent. In contrast, when the AI agent rates a project high, human expert do not blindly follow it. Taken together, this suggests that in the context of high failure rate of the ICOs, human experts use the AI agent as a filter, quickly rejecting projects rated low by the agent, delving deeper into projects rated high. Essentially, we found a combination of algorithm appreciation (Logg et al. 2019) [5] and algorithm aversion (Dietvorst and Bharti 2019) [6] in this context, where the AI agent poses an asymmetric anchoring effect on humans. Finally, our analysis shows that overall, the hybrid human expert and AI system outperforms AI agent alone.

We propose that the AI recommendations have an asymmetric anchoring effect on humans, where humans use AI agent as a filter, and pay more attention to the riskier recommendations (or rare events).

## 2. Prior theory and research

### 2.1. Artificial Intelligence

Artificial Intelligence (AI) is defined as the “ability of a machine to perform cognitive functions that we associate with human minds” (Rai et al. 2019) [7]. A non-exhaustive list of the abilities possessed by AI is the abilities to sense, comprehend, act and learn (Bawack et al. 2019) [8]. The advantage of using AI agents for accomplishing tasks (such as speed, accuracy, reliability, scalability), complement and extend human competencies such as creativity, empathy and judgment (Rai et al. 2019) [7]. The spectrum of involvement of AI in the completion of a task can range from task augmentation to task assemblage, to task substitution (Constantinides 2019) [4]. Firms are beginning to recognize the ability of AI and humans to complement each other, and are beginning to deploy human-AI hybrid processes, which are intelligent systems where of AI agents and human agents work together.

Our focus in this study is on task augmentation. A recent study by Microsoft suggests that 67% of business leaders and 64% of workers believe that AI will augment their work and not displace them, enabling them to do their existing jobs better or reduce repetitive task. (Raisch and Krakowski 2020) [8] argue that organizations adopting a broader perspective comprising both automation and augmentation can achieve complementarities that benefit business and society. However, few studies have investigated the dynamics between human and AI in an AI augmented task context. Our paper is among the first to provide insights on how humans and AI can complement each other in completing complex tasks.

### 2.2. Influence of AI on Decision Making

With the increasing capabilities and applications of AI (Fethi and Pasiouras 2010, Esteva et al. 2017, Leachman and Merlino 2017), firms are beginning to implement AI in decision making. For example, (Edwards et al. 2000) conducted an analysis of AI for business decision making at three organizational decision making levels, i.e. strategic, tactical and operational decisions. Their findings show that: AI can be used to replace human decision makers for structured or semi-structured decisions, but it would be better to be used as a decision support tool for dealing with unstructured decisions at the strategic level in organizations. As the progress of AI technology enables researchers to create advanced machines, it is possible for AI to undertake more complex tasks that require cognitive capabilities such as making tacit judgments, sensing emotion and

driving processes which previously seemed impossible (Mahroof 2019).

Researchers have also presented a range of observations (Logg et al. 2019) regarding how humans react to advice from AI based systems in a decision-making context (Dietvorst and Bharti 2019; Dietvorst et al. 2015; Dietvorst et al. 2018). For instance, recent work by Dietvorst et al. (2018) provides empirical evidence to anecdotal references (Frick 2015; Harrell 2016) which suggest that people choose their own judgment over algorithm forecasts. This phenomenon, termed as algorithm aversion (Dietvorst et al. 2015) embodies human bias against algorithmic advice, despite their established superiority in outperforming human effectiveness (Dawes et al. 1989; Grove et al. 2000).

This aversion to algorithmic advice is particularly salient in contexts when the task outcome is incentivized, such as investing (Dietvorst et al. 2018). Environments which involve high irreducible uncertainty (such as healthcare or medicine) are most likely to elicit algorithm aversion where individual willingness to use algorithms is exacerbated due to the need for outcome precision (Dietvorst and Bharti 2019). However, the ability to modify the outcomes seems to alleviate human aversion to algorithmic advice, highlighting their desire for control over outcomes, particularly in an incentivized context (Dietvorst et al. 2018). Algorithm aversion is also influenced by individuals’ own expertise in decision making, and wanes when people lack expertise, leading to the contrasting phenomenon termed as algorithm appreciation, which motivates individuals to choose algorithmic advice over a human experts’ (Logg et al. 2019). Our study adds to this stream of research by understanding how humans make decisions in volatile markets after being exposed to the decisions made the AI agent. Insights from our study shed light on important considerations for platform designers aiming to use human-AI hybrid models for complex evaluation tasks.

## 3. Theoretical background and hypothesis development

In this study, we examine how assessments made by an AI agent influence the judgments of human experts in their evaluations of Initial Coin Offerings (ICOs). In our context, the platform AI agent assesses and assigns a rating to each ICO as soon as it is published on the platform. This rating is based on an algorithm that takes into consideration observable

and quantitative criteria of the ICO. All experts see the agent rating for an ICO before they rate the ICO.

When AI agent evaluates ICOs, there are generally two types of assessments: either the agent deems that the ICO project does not have the potential to succeed in the long run, and thus provides a low rating, or the agent classifies an ICO project as high quality, and hence gives it a high rating. In the ICO market, there is high likelihood of project failure and fraud, combined with the possibility of gaining extreme returns from a small number of projects. This particular context may lead people to react differently to these two types of evaluations.

We first consider the case where the agent rates a project low. Prospect theory suggests that people are generally loss averse (Kahneman and Tversky 2013)—individuals assess their loss and gain perspectives in an asymmetric manner, the pain of loss a certain amount of money is more pronounced than the pleasure of gaining the same amount of money. For example, for some individuals, the pain from losing \$1000 could only be compensated by the pleasure of earning \$2000. Therefore, in a highly risky market where the probability of gain is very low, and investments in most projects will end up being losses, experts are likely to be more risk averse and put high weight on low AI agent ratings. In this situation, AI agent's low rating may have an "anchoring effect" on the experts (Strack and Mussweiler 1997; Tversky and Kahneman 1974), conditioning them to focus on the potential loss of the investment.

The task of evaluating an ICO is complex. The true potential of an ICO depends on a myriad of factors and is extremely difficult to gauge, even for domain experts. ICO teams know significantly more about their offering than evaluators, inducing information asymmetry. The ICO platform has designed the AI agent to provide an assessment of each ICO based on an array of observable information. The intent is that the AI agent will offer an expert opinion that is unbiased. Of course, the accuracy of the agent's assessment is an empirical question, but it is advertised as a reliable resource to facilitate the human experts and investors. Therefore, a risk averse human expert who has access to the agent's assessment should integrate its low rating into their own analysis processes. Banerjee explains that there is a rationale for agents to take considerations of prior agents' evaluations, because these other decision makers may have some information that the agents do not have (Banerjee 1992). In our context for example, if human experts believe that the AI agent is reasonably good at collecting information, then the human experts are likely to be influenced by the

agent. Thus, it is likely that the risk averse experts will put more weight on AI agent's low rating and treat it as a signal of potential risk and give a low rating following the AI agent. Therefore:

***H1: When an AI agent provides a low rating to a project, subsequent human ratings are positively related to the AI agent's ratings.***

Consider the case in which the agent gives a high rating. As discussed earlier, the high failure rate in the ICO market is common knowledge in the community (Rhue 2018), thus investors and experts are likely to be risk averse. Also well-known is the fact that the AI agent can only incorporate objective and observable criteria into its assessment. Taken together, these two reasons increase humans' skepticism of high agent ratings. This skepticism steers human experts away from the inclination to passively align with the agent's judgment when the agent suggests an ICO project is worthy of investment, instead the skepticism motivates human experts to analyze such projects critically. Human experts on the platform have limited time and resources, and they choose to invest these resources into analyzing the projects that pass the agent test. In other words, humans may tend to use the agent as a filter and consider only projects that pass through this filter worthy of detailed and critical examination.

When applied carefully, human judgment is valuable because of the ability of humans to reason in a way that goes beyond executing rote calculations on available data (McAfee and Bynjolfsson 2017). As opposed to "explicit knowledge", which is formal, codified and can be readily explained, humans possess "tacit knowledge", the kind of knowledge we are often not aware we have, and is therefore difficult to transfer to a machine (Polanyi 1958). "We can know more than we can tell", which came to be known as Polanyi's paradox (Polanyi 1966), captures the fact that we tacitly know a lot about the way the world works, yet aren't able to explicitly describe this knowledge. This tacit knowledge of human experts in the ICO market allows them to take into consideration several intangible aspects of the project such as the vision, innovativeness and potential, which are not accounted for in the agent's evaluation. We therefore expect that humans will critically examine these projects and agree with the agent only when they believe that the agent's evaluation is correct even after applying their own tacit knowledge. Critical examination will enable humans to detect spurious cases of high agent ratings and they are likely to disagree with the agent in these cases. Therefore:

***H2. When an AI agent provides a low rating to a project, subsequent human ratings are not related to the AI agent's ratings.***

Our context represents a particular type of system where humans are exposed to AI outputs before making decisions. Such systems have wide applications, such as helping physicians make diagnosis, or recommending product and services in e-commerce. In such systems, humans are aware of the odds of an event happening, in our case, an ICO project succeed in the long run, which is low. Therefore, a rational human would infer this is a high-risk context and one should be more cautious. Thus, when AI recommends investing in a project, careful evaluations based not only on observable information, but also based on tacit knowledge is needed. In contrast, when AI agent rates a project poorly, it is very possible that the project does not meet certain basic criteria. Given that AI is more efficient in collecting and analyzing objective information than human, the low rating would be a worth considering signal of low quality. Hence, in such systems, it is possible for humans to combine the strengths of both human and AI intelligence to achieve a better overall decision making. Therefore, we hypothesize:

***H3: Humans AI augmented systems can make better decisions than AI agents alone.***

Another interesting question that stems from our hypotheses is when and why humans deviate from the AI agent in their evaluations? In other words, what do human experts focus on or take into consideration when they make independent evaluations? Do these factors depend on whether the AI agent's evaluation was favorable or unfavorable? And more importantly, when do these considerations help them make better decisions? We explore these questions in post-hoc analyses by taking a deeper look at the content of the reviews written by human experts.

## **4. Method**

### **4.1. Research Context**

Following related research in the ICO context (Bourveau et al. 2018), our data is collected from ICObench.com, which is one of the most prominent and comprehensive ICO reviewing and rating platform (Note that, the actual process of fund raising is not performed in ICObench.com or other rating sites). On ICObench.com, ICO teams promote their projects with a pitching video, white paper, descriptions of the projects and other related financial and team member information. Once a project is

posted, the AI agent of the platform will provide a rating for the project. Then, a community of experts (people who are experienced and active in the ICO community) will provide reviews and ratings for these projects. Using these insights, investors can choose promising ICOs to invest.

The platform displays the data of each ICO on three tabs. First, on the main page, ICO project related information is provided comprising a brief introduction, the project pitch, and the goals of the project. It also includes other details such as the token type, value of token in ICO, ICO soft cap, hard cap. The AI agent rating is presented on this main tab. Second, the ratings tab displays the experts' reviews and ratings of the ICO and links to the profile page of each expert who rated the ICO. Third, the Team tab lists the profiles of the ICO team members.

Each ICO project receives two types of ratings, one by the AI agent of this platform, popularly known as Benchy, and the second by experts, who are active members of the ICO Bench community and experienced domain experts. The Agent uses an algorithm taking into inputs of a number of objective and observable parameters to evaluate each of the ICO projects. The agent evaluates each ICO on 4 major dimensions, namely, Team, ICO Information, Product Presentation and Marketing & Social Media. The human experts on the other hand provide a more subjective and qualitative evaluation. Each human expert provides a rating on a scale of 1 to 5 on three broad categories: Team, Vision and Product. Apart from the ratings the human experts also provide a textual review explaining their evaluations.

### **4.2. Data**

We collected data of 2783 ICO projects, from January 2017 to December 2018. For each project, we collected individual expert ratings and reviews, the date the rating was provided, as well as ICO project data. Some of our dependent, independent and control variables were available in the data set, while others were constructed for the analysis. The descriptive statistics for these variables are presented in table 1.

#### **Dependent Variable**

Expert Rating is an average rating on a scale of 1 to 5 given to an ICO project by each individual human expert based on three primary parameters of team, vision and product. It has a mean of 3.79 and a standard deviation of 0.584 and the distribution of the ratings is positively skewed. The distribution of the expert rating is presented in figure 1.

#### **Independent Variable**

Agent Rating is the evaluation provided at the beginning by the artificial agent to each of the ICO projects. It has a mean of 3.71 and a standard deviation of 0.632 showing the agent is more conservative in its overall evaluations. The distribution of the agent rating is presented in figure 2. Figure 3 shows the distribution of the difference between expert ratings and agent rating.

**Control Variables**

PreICO is a binary variable, which indicates whether the projects have a Pre-ICO fund raising period prior to the official ICO.

Price in ICO is measured as the value of each token in US Dollars sold during the ICO stage.

Distributed in ICO is measured as the percentage of the token that will be distributed during the ICO.

Softcap is the minimum amount of funding an ICO project team aims to raise during the ICO. Here Softcap is a categorical variable with the values of 0, 1 or 2, where if the softcap information is not available on the icobench website, the whitepaper and the project website, then it is coded as 0. If the value of softcap is between 0 and 50 percentile then it is coded as 1, and if the softcap is higher than 50 percentile, it is coded as 2.

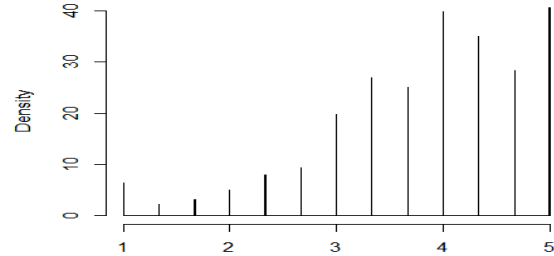
Hardcap is the maximum amount of funding an ICO team will accept during the ICO. Once the hard cap is reached, no future funding from investors will be accepted. Hardcap is a categorical variable, defined as 0, 1 or 2, where if the hardcap information is not available on the icobench website, the whitepaper and the project website, then it is coded as 0. If the value of softcap is below the 50<sup>th</sup> percentile, it is coded as 1, and 2 otherwise.

Success is a proxy for the quality of the project and potential to be successful. Prior research has used whether a token sold during ICO was successfully listed on exchanges after the ICO, and whether the ICO achieved its softcap, and amount raised in ICO to measure the success of ICO projects (Bourveau et al. 2018; Lee et al. 2019; Li et al. 2019; Lyandres et al. 2019). We use whether the ICO token is trading at a non-zero value one year after the ICO (considering if a token is successfully listed in an exchange and continued trading for one year) to measure the ICO project success.

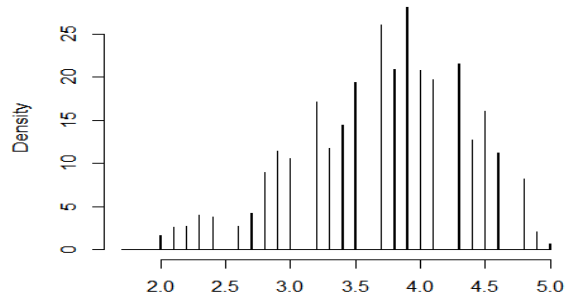
**Table 1. Summary Statistics**

Variable	Mean	Std. Dev.	Min	Max
Expert Rating	3.79	0.584	0	5
Agent Rating	3.71	0.632	1.7	5
PreICO	0.65	0.477	0	1
Price in ICO	0.134	0.205	0	2.762

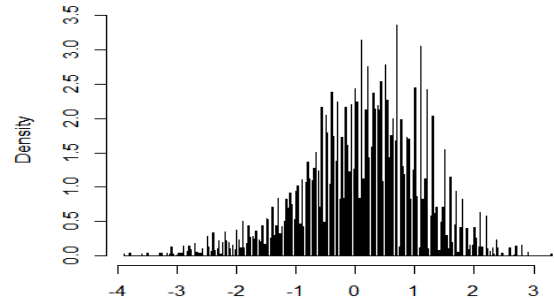
Distributed in ICO	0.418	0.297	0	1
Softcap	0.637	0.673	0	2
Hardcap	1.231	0.844	0	2
Success	0.179	0.383	0	1



**Figure 1. Histogram of expert ratings**



**Figure 2. Histogram of agent ratings**



**Figure 3. Histogram of (Expert Ratings – Agent Rating)**

**5. Analysis**

In Table 2, we share our preliminary analysis where we regress the expert rating on agent rating. The results reveal a positive relationship between agent ratings and expert ratings while controlling for ICO project characteristics (0.437, p<0.001).

**Table 2. Linear Regression of effect of agent rating on expert ratings**

	ExpertRating
Agent Rating	0.437*** (0.014)
PreICO	0.068*** (0.019)
Price in ICO	-0.016

	(0.044)
<i>Distributed in ICO</i>	0.101***
	(0.030)
<i>Soft Cap</i>	0.018
	(0.014)
<i>Hard Cap</i>	0.012
	(0.011)
<i>Intercept</i>	2.117***
	(0.058)

Note: \*  $p < 0.05$ ; \*\*  $p < 0.01$ ; \*\*\*  $p < 0.001$

However, the positive relationship between agent ratings and expert ratings could simply mean that both the bots and the experts happen to be correct (or incorrect) in their respective judgments of ICOs due to unobserved variables. The regression above does not allow us to make casual inferences about the effect that agent ratings have on expert ratings. Below, we share the challenges and our empirical strategies in examining whether agent ratings influence expert ratings. In particular, following prior literature, we use the system GMM model to address the issue of endogeneity of our dependent variable, *Expert Rating*, by using transformed regressors as valid instruments (Ghose and Han 2011, Li and Wu 2018, Aral, Brynjolfsson and Van Alstyne 2012, Yoganarasimhan 2012, and Sonier, McAlister and Rutz 2011).

### 5.2. Econometric Model

We developed a dynamic model of expert ratings of ICOs. Our model is dynamic because we account for the fact that our y variable, namely, expert rating, depends on its lagged values, i.e., mean of aggregate ratings from previous periods. Such state-dependence is important to consider while studying outcomes that build on an “accumulated stock” over time. Recent examples of such phenomena studied using dynamic models include mobile-user content consumption (Ghose and Han 2011), sales (Li and Wu 2018), project output (Aral, Brynjolfsson and Van Alstyne 2012), YouTube video views (Yoganarasimhan 2012), and firm sales (Sonier, McAlister and Rutz 2011). Our empirical specification is presented in the following model, where

$$y_{ij} = c + \alpha \frac{1}{n} \sum_{n=1}^{j-1} (y_{in}) + \beta X_i + \delta_i + \varepsilon_{ij}$$

$y_{ij}$  denotes the expert  $j$ 's rating for project  $i$ ,  $Y_{ij-1}$  is the lagged average of the dependent variable, i.e., average rating by all experts before expert  $j$  for project  $i$ ,  $X_i$  is a vector of project-specific expert-invariant covariates. It includes our explanatory variable, agent rating. In addition, it also has all the control variables, namely, *Preico*, *Price*, *DistributedInICO*, *Softcap* and *Hardcap*, and  $\delta_i$  is an expert-invariant unobserved project fixed effect. It captures other inherent characteristics of the ICO.  $\varepsilon_{ij}$

is the idiosyncratic error term that captures random shocks to the ICO evaluations.

### 5.3. Estimation Strategy

The dynamic model used here for panel data analysis is System General Method of Moments (GMM) and this enables the explanatory variables to be treated as potentially endogenous. It is not difficult to see that most of our expert-varying variables  $X_i$  in this analysis can be potentially endogenous and can be correlated with ICOs' fixed or expert-varying unobserved factors. An initial remedy to this issue might be to use the fixed-effect models. However, Nickell (1981) show that in a dynamic setting with finite T, fixed-effect estimators are inconsistent. To resolve this issue, a method of using instrumental variables (IVs) is suggested by Anderson and Hsiao (1981) and further developed by Arellano and Bond (1991). Their method which is also called “difference GMM,” is suitable for “small T, large N” panels. Assuming that  $\varepsilon_{ij}$ 's are iid across  $i$  and  $j$ , the Arellano-Bond estimation starts by transforming all regressors, usually by “first-differences” or “forward orthogonal deviations” to eliminate fixed effect  $\eta_i$ . Then they show that longer lags of the regressors can be used as valid instruments, and applying the Generalized Method of Moments (GMM) provides a consistent and efficient estimator. However, Blundell and Bond (1998) show that in dynamic panel models where the autoregressive parameter ( $\alpha$ ) is moderately large and the number of time series observations is moderately small, the Arellano-Bond IVs (past levels of the regressors) convey little information about the transformed regressors (future changes) and therefore the weak instruments make the Arellano-Bond estimator perform poorly. By assuming an initial condition, Blundell and Bond (1998) shows that in a level (untransformed) equation one can use transformed regressors as valid instruments orthogonal to the fixed effects. Adding the new moment conditions to the set of Arellano-Bond moment conditions, they designed an efficient “system GMM” estimator that performs better than “difference GMM.” To run the system GMM model, we use `xtabond2`, a stata command written by Roodman (2006). With this command we use a two-step option to make analysis robust to heteroscedasticity. Further, we use the robust option to apply the Windmeijer (2005) finite-sample correction to fix the downward bias of the system GMM standard errors. Roodman (2009) pointed out that too many instruments in system GMM models can result in over-fitting the endogenous variables. Therefore, we use the collapse option which creates one instrument for each variable and lag distance, rather than one for each time period, variable, and lag

distance. This approach in small samples can avoid the bias due to the rising number of instruments. Finally, we add the orthogonal option which requests the forward-orthogonal-deviations transformation instead of first differencing. We examine and show the validity of the group of IVs by using Hansen test and we have also recorded the AR (2) p-value to further validate our results and implied consistency across all the models.

#### 5.4. Model Specifications

Model 1 is our baseline model. We run the model described above on the entire data set (Sample 1 in Figure X). For models 2 and 3, we use the same specification but run the models on different sub samples. In Model 2, we examine only instances where agent rating is low and in Model 3, only those observations where the agent rating is high (Sample 3). While it might seem like an option to do an interaction analysis instead of sub-sample analysis, note that we are examining the effect of only one independent variable, AI Agent Rating, for low and high values of the same variable.

We classify agent ratings greater than 4.3 as “high” and the rest as “low”. The rationale for this choice is as follows. In our sample, 15% of all projects were successful (following prior literature(Amsden and Schweizer 2018; Momtaz 2020),we classify a project as successful if the ICO token is trading at a non-zero value one year after the ICO). Hence based on the agent rating distribution, we choose the top 15<sup>th</sup> percentile (i.e. 4.3 and above) of the AI agent ratings as the cut off for successful rating. We repeat our analysis with different thresholds (4.0 and 4.6) and find that our results remain qualitatively consistent throughout.

#### 5.5. Results

In Model 1 we estimate the effect of agent rating on expert rating using GMM estimation, controlling for lagged *expert rating*, *preICO*, *Price in ICO*, *Distributed in ICO*, *softcap*, *hardcap* and *success*. We see agent rating has a positive and statistically significant effect on expert rating (0.473,  $p < 0.01$ ). In model 2, which is the sample with low agent ratings, we observe that the agent ratings have a positive and statistically significant effect on expert rating (1.136,  $p < 0.01$ ). This shows that when AI agent rates a project low, experts are likely to be aligning with the agent.

However, in model 3, which is the sample with high agent ratings, we do not find a statistically significant effect of the agent rating on expert ratings ( $p = 0.322$ ). This shows that when the agent is positive about the likelihood of success of an ICO, experts are possibly

more cautious and wary of aligning with the agent’s assessment.

In our main analysis, we choose 4.3 as the cut off because it is representative of the success likelihood in our sample.

**Table 3. System GMM Estimation of the Effect of Agent Ratings on Expert Ratings**

	M1: Full Sample	M2: Agent Low	M3: Agent High
<i>Agent Rating</i>	<b>0.4727*</b> <b>(0.2826)</b>	<b>1.1356***</b> <b>(0.4306)</b>	<b>-2.0199</b> <b>(2.0393)</b>
<i>Lagged Rating</i>	0.2747 (0.3473)	0.3521 (0.3133)	-0.2803 (0.5315)
<i>Pre ICO</i>	-0.1197 (0.2903)	-0.3517 (0.4389)	0.4114 (0.4715)
<i>Price in ICO</i>	-1.1948 (1.0552)	-1.2546 (1.0264)	3.4311 (4.1724)
<i>Distributed in ICO</i>	0.2049 (0.8876)	-0.9296 (0.7861)	0.3646 (0.8203)
<i>Softcap</i>	0.2309 (0.2007)	0.2478 (0.1884)	-0.4635 (6.4826)
<i>Hardcap</i>	-0.2215 (0.3082)	-0.2247 (0.3064)	-0.3255 (0.3164)
<i>Success</i>	0.3246 (0.6183)	0.3667 (0.5539)	-0.1501 (0.8189)
<i>Intercept</i>	1.2699 (1.7895)	-0.6952 (1.2453)	14.3155 (10.3179)
No. of Observations	9339	6891	2448
AR(2) p-value	0.464	0.570	0.700

Note: \*  $p < 0.1$ ; \*\*  $p < 0.05$ ; \*\*\*  $p < 0.01$

#### 5.6. Robustness Checks

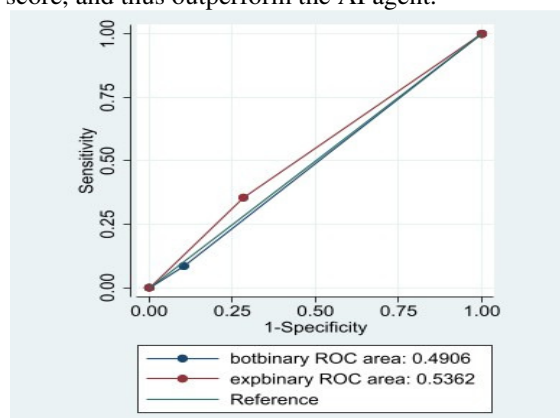
In order to verify the robustness of our results, we alter the threshold for “high rating” by the bot. In our main analysis, we choose 4.3 as the cut off because it is representative of the success likelihood in our sample. Adopting this procedure helps us to prove the trends remain consistent over a considerable range showing that the relation between humans and the AI agent hold generally. Using 4.0 and 4.6 as the benchmark agent ratings, we found that the results are qualitatively similar to the main models. We found that in both models, the AI rating has statistically significant effect on expert ratings for samples 1 (full sample), 2 (AI rates low), but not in model 3 (AI rates high). We also used three different dependent variables. We used the individual ratings provided by the Human Experts on each category of Team, Vision and Product (instead of the aggregate rating, which we used in our main analysis). Here, we show that even for each of the category-wise ratings, the results are qualitatively similar. Due to space constraints, we do not present these results.

#### 5.7. Evaluating Human and AI Performances



To test hypothesis 4, we perform the following analysis to compare human experts and AI agent's performances. Based on the confusion matrix, we calculate the precision, recall, F score, as well as sensitivity and specificity for both AI and human experts. We also plot the ROC curve to compare the performances of AI and human experts.

Although the AI agent has a high accuracy than human experts, we need to consider the context of the analyses. The fact that more than 80% of ICOs have historically failed, means that an AI agent which always predicts failure will be correct more than 80% of the time. In such situations, the two more meaningful model evaluation metrics are Precision and Recall. In our case, Precision indicates what proportion of ICOs rated as high quality actually succeeds. On the other hand, Recall indicates the percentage of total successful projects that are correctly rated high. The F1-score is a harmonic mean of the precision and recall which balances the use of precision and recall to measure performance. Human experts have higher precision, recall, and F1 score, and thus outperform the AI agent.



**Figure 4. ROC/ AUC Curve for Human and AI Agents**

In addition, we also evaluated the sensitivity and specificity of human experts and the AI agent. Sensitivity indicates what proportion of the positive class is correctly classified, i.e. what proportion of successful ICOs is correctly predicted. Specificity indicates what proportion of the negative class are correctly classified, i.e. how accurate human and AI are in predicting the failed ICOs. These two metrics are graphically represented using a Receiver Operator Characteristic (ROC) curve for an easier understanding of the prediction performances. The ROC curve is a probability curve that plots the True Positive Rate against False Positive Rate at various threshold values. TPR is same as sensitivity/recall while FPR is (1-specificity) or the proportion of the negative class that got incorrectly classified. The

Area Under the Curve (AUC) is the measure of the ability of a classifier to distinguish between classes and is used as a summary of the ROC curve. The higher the AUC value the better is the prediction performance. Based on ROC/AUC analysis, we can see that the agent performs about at the level of random chance, and the human experts perform only slightly better.

## 6. Discussion

Our study provides useful findings on human behaviors in settings where humans perform a complex task incorporating advice from an AI agent. We find that humans' ratings of initial coin offerings outperformed the ratings of the AI agent alone, and strike a better balance between precision and recall. We noted that neither the human experts nor the agent had very good performance in our particular context. Human ratings were influenced by the ratings given by the AI agent, but in an asymmetrical manner. In our context, 80% of the projects failed, so as we theorized, human experts' decisions were aligned with that of the AI agent's when the agent rated a project low, but were not aligned when the agent rated a project high. Thus, we conclude that most experts used the agent's assessments to screen out projects and usually concurred with the agent when it gave low ratings. When an agent gave high ratings to a project, the human experts focused more carefully on it, and made judgments independent of the agent's assessment. In other words, a high agent rating triggered experts to apply their own judgment, which was not materially influenced by the agent's ratings. We suggest that in high risk contexts, AI poses an asymmetric anchoring impact on humans.

There are several reasons for the asymmetric anchoring we observe. First, predicting success is a much riskier bet than predicting failure in the ICO market with its high failure rate. Second, it is possible that the humans are using the agent as a filter or a first check. When the agent rates low, humans interpret that as the project not even meeting a threshold and may not give it full consideration. But, when the agent rates high, it passes the first check, and experts give these projects due attention, after which they agree or disagree with the agent.

The tendency of human experts to align their ratings with those of the AI agent has to do with the nature of the task at hand. The true potential of an ICO is extremely difficult to evaluate, and when faced with such problems that are highly complex and uncertain in nature, it is likely that humans lean on objective signals. It is important for the designers



of human – AI hybrid systems and platforms to bear this in mind and carefully consider when, whether, and to whom the work of the agent should be made visible.

We use three different measures of prediction quality to compare the performances of the agent and the human and to analyze if adding a human to the mix adds value. Overall, the AI agent has a higher accuracy than the human in its predictions, but this is largely driven by the fact the agent is more conservative in its assessments. We then find that humans perform better at both precision and recall, thus a higher F1 score overall. Humans are also better at distinguishing between the two classes, namely high-quality and low-quality projects, as depicted in the ROC curve in Figure 4.

Our study has several limitations. First, we use whether the ICO token is trading at a non-zero value one year after the ICO as the measure for success. We choose this measure based on prior research (Amsden and Schweizer 2018; Momtaz 2020) but we recognize that success is multi-dimensional and there are several other ways to measure it. Second, our understanding of *Benchy*, the platform agent's algorithm is limited. While we are aware of some of the inputs to the algorithm based on the information on icobench.com, we don't know the details of the algorithm, and are unable to factor it into our analysis. Finally, we acknowledge the limitation on the generalizability of our results which like all empirical results are contingent on the characteristics of the problem at hand, the context, and the market we examine.

We believe the findings offer three implications for theory and future research. First, our results show that in this form of human-AI partnership, human experts rely on the AI agent to screen out projects and focus their attention on projects the AI has rated highly. Our findings are influenced by our research context which is characterized by (a) the complexity of the task at hand (b) the volatile nature of ICOs and the risks involved, (c) the skewness of the outcomes – the algorithm is used to identify rare events, and (d) the sequential nature of task completion (the AI agent completes the task first and the human experts use the AI agent's assessment in their own judgment). Future research can verify this finding in similar contexts, such as diagnose of rare diseases, and identifying investment opportunities.

Second, our study opens avenues for future research on the human–AI partnership. It is well-known that human bias can creep into algorithms (Chan and Wang 2017, Dastin 2018), and humans exhibit bias for or against AI algorithms (Dietvorst et al. 2015, Luo et. al 2019, Al-Natour et al. 2006,

Benbasat and Wang 2005). Adopting prospect theory into human AI interaction context, our study shows how humans weigh positive and negative AI recommendations differently due to risk aversion tendencies. This brings up interesting questions, when designing AI augmented systems, should we take human cognitive biases into consideration? How to best design systems for different contexts?

Third, our study builds on research in the human and Decision Support Systems interaction area and investigates human behavior when interacting with Intelligent Decision Support Systems. Unlike prior research suggests AI algorithm appreciation or aversion in different contexts, our study finds an asymmetric anchoring effect of the AI agents on human in contexts with high risk and uncertainty.

Our findings also have three implications for users and designers of human–AI hybrid systems. First, system designers need to carefully contemplate on the desired roles of the agent and the human. How important is the independence of human judgment in the context? Is the intended role of the agent that of an assistant that makes the human's job easier via a preliminary pre-screening analysis? Or are the agent and human meant to act as independent, complementary undertakers of the task, given the comparative strengths of AI agents and humans? Depending on the intended role, system designers need to decide whether to expose the human to the work product of the agent.

Second, if humans are inclined to use the AI agent as a filter for screening, there may be value in designing AI agents that do just that. In other words, instead of aiming to have the AI do the same job as the human, we could potentially get better results if we focused on creating AI agents that weed out distinctly poor prospects and create shortlists for humans to examine carefully.

Finally, our study shows that we cannot put humans and AI together to solve complex problems and expect superior performance to naturally emerge. From a team-building perspective, just as we train team members of complementary strengths to work with each other effectively, in a world with increasing integration between human and artificial intelligences, we see the need for the deliberate and systematic training of humans to work effectively alongside AI counterparts.

## 7. References

[1] Fethi, M.D. and Pasiouras, F., "Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey," *European journal of operational research*, 204(2), 2010, pp.189-198.

- [2] Esteva, A., Kuprel, B., Novoa, R.A., Ko, J., Swetter, S.M., Blau, H.M. and Thrun, S., "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, 542(7639), 2017, pp.115-118.
- [3] Leachman, S.A. and Merlino, G., "Medicine: The final frontier in cancer diagnosis," *Nature*, 542(7639), 2017, pp.36-38.
- [4] Constantinides, P., "How Human-Ai Hybrids Will Change Work Forever," *Warwick Business School*, 2019.
- [5] Rai, A., Constantinides, P. and Sarker, S., "Editor'S comments: next-generation digital platforms: toward human-AI hybrids," *Mis Quarterly*, 43(1), 2019, pp.iii-x.
- [6] Bawack, R., Wamba, S.F. and Carillo, K., "Where Information Systems Research Meets Artificial Intelligence Practice: Towards the Development of an AI Capability Framework," *Technology*, 12, 2019, pp.15-2019.
- [7] Raisch, S. and Krakowski, S., "Artificial Intelligence and Management: The Automation-Augmentation Paradox," *Academy of Management Review*, (ja), 2020.
- [8] Edwards, J.S., Duan, Y. and Robins, P.C., "An analysis of expert systems for business decision making at different levels and in different roles", *European Journal of Information Systems*, 9(1), 2000, pp.36-46.
- [9] Mahroof, K., "A human-centric perspective exploring the readiness towards smart warehousing: The case of a large retail distribution warehouse," *International Journal of Information Management*, 45, 2019, pp.176-190.
- [10] Logg, J.M., Minson, J.A. and Moore, D.A., "Algorithm appreciation: People prefer algorithmic to human judgment," *Organizational Behavior and Human Decision Processes*, 151, 2019, pp.90-103.
- [11] Dietvorst, B. and Bharti, S., "People Reject Even the Best Possible Algorithm in Uncertain Decision Domains," *Available at SSRN 3424158*, 2019.
- [12] Dietvorst, B. J., Simmons, J. P., and Massey, C., "Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them," *Management Science* (64:3), 2018, pp. 1155-1170.
- [13] Harrell, E., "Managers shouldn't fear algorithm-based decision making," *Harvard Business Review*, 2016.
- McAfee, A. and Brynjolfsson, E., "*Machine, platform, crowd: Harnessing our digital future*," WW Norton & Company, 2017.
- [14] Banerjee, A. V., "A Simple Model of Herd Behavior," *The quarterly journal of economics* (107:3), 1992, pp. 797-817.
- [15] Bikhchandani, S., Hirshleifer, D., and Welch, I., "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of political Economy* (100:5), 1992, pp. 992-1026.
- [16] Herzenstein, M., Dholakia, U. M., and Andrews, R. L., "Strategic Herding Behavior in Peer-to-Peer Loan Auctions," *Journal of Interactive Marketing* (25:1), 2011, pp. 27-36.
- [17] Zhang, J. and Liu, P., "Rational herding in microloan markets," *Management science*, 58(5), 2012, pp.892-912.
- [18] Agrawal, A.K., Catalini, C. and Goldfarb, A., "*The geography of crowdfunding* (No. w16820)," National bureau of economic research, 2011.
- [19] Kahneman, D. and Tversky, A., "Prospect theory: An analysis of decision under risk," In *Handbook of the fundamentals of financial decision making*, 2013 (pp. 99-127).
- [20] March, J. G., and Shapira, Z., "Managerial Perspectives on Risk and Risk Taking," *Management science* (33:11), 1987, pp. 1404-1418.
- [21] Thaler, R. H., and Johnson, E. J., "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice," *Management science* (36:6), 1990, pp. 643-660.
- [22] Tversky, A., and Kahneman, D., "Judgment under Uncertainty: Heuristics and Biases," *science* (185:4157), 1974, pp. 1124-1131.
- [23] Shafir, E., and LeBoeuf, R. A., "Rationality," *Annual review of psychology* (53:1), 2002, pp. 491-517.
- [24] McAfee, A. and Brynjolfsson, E., "*Machine, platform, crowd: Harnessing our digital future*," WW Norton & Company, 2017.
- [25] Rhue, L., "Trust is all you need: an empirical exploration of initial coin offerings (ICOs) and ICO reputation scores," *Available at SSRN 3179723*, 2018.
- [26] Bourveau, T., De George, E.T., Ellahie, A. and Macciocchi, D., "Initial coin offerings: Early evidence on the role of disclosure in the unregulated crypto market," *Available at SSRN 3193392*, 2018.
- [27] Aral, S., Brynjolfsson, E. and Van Alstyne, M., "Information, technology, and information worker productivity," *Information Systems Research*, 23(3-part-2), 2012, pp.849-867.
- [28] Anderson, T.W. and Hsiao, C., "Estimation of dynamic models with error components," *Journal of the American statistical Association*, 76(375), 1981, pp.598-606.
- [29] Arellano, M. and Bond, S., "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The review of economic studies*, 58(2), 1991, pp.277-297.
- [30] Blundell, R. and Bond, S., "Initial conditions and moment restrictions in dynamic panel data models," *Journal of econometrics*, 87(1), 1998, pp.115-143.
- [31] Roodman, D., "How to do xtabond2: An introduction to. *Difference*" and "*System*" *GMM in Stata*", *Center for Global Development Working Paper*, 103, 2006.
- [32] Windmeijer, F., "A finite sample correction for the variance of linear efficient two-step GMM estimators," *Journal of econometrics*, 126(1), 2005, pp.25-51.
- [33] Lee, J., Li, T., and Shin, D., "The Wisdom of Crowds in Fintech: Evidence from Initial Coin Offerings," *Available at SSRN 3195877*, 2019.
- [34] Liang, T.-P., "Recommendation Systems for Decision Support: An Editorial Introduction." Elsevier, 2008.