Pharmacists' Trust in Automated Pill Recognition Technology: The Role of Presenting AI Uncertainty Information

Jin Yong Kim, Vincent D. Marshall, Brigid Rowell, Qiyuan Chen, Yifan Zheng, John D. Lee, Raed Al Kontar, Corey Lester, X. Jessie Yang

Submitted to: Journal of Medical Internet Research
on: February 28, 2024

Table of Contents

Pharmacists' Trust in Automated Pill Recognition Technology: The Role of Presenting AI Uncertainty Information

Jin Yong Kim[1] BSE; Vincent D. Marshall[2] MS; Brigid Rowell[2] MA; Qiyuan Chen[1] MSE; Yifan Zheng[2] PharmD; John D. Lee[3] PhD; Raed Al Kontar[1] PhD; Corey Lester[2] PhD, PharmD; X. Jessie Yang[1] PhD

[1]Industrial and Operations Engineering University of Michigan Ann Arbor US
[2]University of Michigan College of Pharmacy Ann Arbor US
[3]Industrial and Systems Engineering University of Wisconsin-Madison Madison US

Corresponding Author:
X. Jessie Yang PhD
Industrial and Operations Engineering
University of Michigan
1205 Beal Ave.
Ann Arbor
US

Abstract

Background: Dispensing errors significantly contribute to adverse drug events, resulting in substantial healthcare costs and patient harm. Automated pill verification technologies have been developed to aid pharmacists with medication dispensing. However, pharmacists' trust in such automated technologies remains unexplored.

Objective: This study aims to investigate pharmacists' trust in automated pill verification technology designed to support medication dispensing.

Methods: Thirty participants performed a simulated pill verification task with the help of an imperfect AI aid recommending acceptance or rejection of a filled medication. The experiment employed a mixed-subjects design. The between-subjects factor was the AI aid type, with or without an AI uncertainty plot. The within-subjects factor was the four potential verification outcomes: AI rejects the incorrect drug, AI rejects the correct drug, AI approves the incorrect drug, and AI approves the correct drug.

Results: Participants had an average trust propensity score of 72 out of 100 (SD = 18.08), indicating a positive attitude towards trusting automated technologies. The introduction of an uncertainty plot to the AI aid significantly enhanced pharmacists' end trust ($t(28) = -1.854$, $p = .037$). Trust dynamics were influenced by AI aid type and verification outcome. Specifically, pharmacists using the AI aid with the uncertainty plot had a significantly larger trust increment when AI approved the correct drug ($t(78.98) = 3.93$, $p < .001$) and a significantly larger trust decrement when AI approved the incorrect drug ($t(2939.72) = -4.78$, $p < .001$). Intriguingly, the absence of the uncertainty plot led to an increase in trust when AI correctly rejected an incorrect drug, whereas the presence of the plot resulted in a decrease in trust under the same circumstances ($t(509.77) = -3.96$, $p < .001$). A pronounced "negativity bias" was observed, where the degree of trust reduction when AI made an error exceeded the trust gain when AI made a correct decision ($z = -11.30$, $p < .001$).

Conclusions: To the best of our knowledge, our study is the first attempt to examine pharmacists' trust in automated pill verification technology. Our findings reveal that pharmacists have a favorable disposition toward trusting automation. Moreover, providing uncertainty information about the AI's recommendation significantly boosts pharmacists' trust in the AI aid, highlighting the importance of developing transparent AI systems within healthcare.

(JMIR Preprints 28/02/2024:57835)
DOI: https://doi.org/10.2196/preprints.57835

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Pharmacists' Trust in Automated Pill Recognition Technology: The Role of Presenting AI Uncertainty Information

Authors: Jin Yong Kim[1], Vincent Marshall[2], Brigid Rowell[2], Qiyuan Chen[1], Yifan Zheng[2], John D. Lee[3], Raed Al Kontar[1], Corey Lester[2], X. Jessie Yang[1]

[1]Industrial and Operations Engineering, University of Michigan

[2]College of Pharmacy, University of Michigan

[3]Industrial and Systems Engineering, University of Wisconsin-Madison

Corresponding Author: Dr. X. Jessie Yang

Email: xijyang@umich.edu

Mailing Address: 1205 Beal Avenue, Ann Arbor, MI, 48109

Phone: 734-763-0541

ABSTRACT

Background: Dispensing errors significantly contribute to adverse drug events, resulting in substantial healthcare costs and patient harm. Automated pill verification technologies have been developed to aid pharmacists with medication dispensing. However, pharmacists' trust in such automated technologies remains unexplored.

Objective: This study aims to investigate pharmacists' trust in automated pill verification technology designed to support medication dispensing.

Methods: Thirty participants performed a simulated pill verification task with the help of an imperfect AI aid recommending acceptance or rejection of a filled medication. The experiment employed a mixed-subjects design. The between-subjects factor was the AI aid type, with or without an AI uncertainty plot. The within-subjects factor was the four potential verification outcomes: AI rejects the incorrect drug, AI rejects the correct drug, AI approves the incorrect drug, and AI approves the correct drug.

Results: Participants had an average trust propensity score of 72 out of 100 (SD = 18.08), indicating a positive attitude towards trusting automated technologies. The introduction of an uncertainty plot to the AI aid significantly enhanced pharmacists' end trust ($t(28) = -1.854$, $p = .037$). Trust dynamics were influenced by AI aid type and verification outcome. Specifically, pharmacists using the AI aid with the uncertainty plot had a significantly larger trust increment when AI approved the correct drug ($t(78.98) = 3.93$, $p < .001$) and a significantly larger trust decrement when AI approved the incorrect drug ($t(2939.72) = -4.78$, $p < .001$). Intriguingly, the absence of the uncertainty plot led to an increase in trust when AI correctly rejected an incorrect drug, whereas the presence of the plot resulted in a decrease in trust under the same circumstances ($t(509.77) = -3.96$, $p < .001$). A pronounced "negativity bias" was observed, where the degree of trust reduction when AI made an error exceeded the trust gain when AI made a correct decision ($z = -11.30$, $p < .001$).

Conclusions: To the best of our knowledge, our study is the first attempt to examine pharmacists' trust in automated pill verification technology. Our findings reveal that pharmacists have a favorable disposition toward trusting automation. Moreover, providing uncertainty information about the AI's recommendation significantly boosts pharmacists' trust in the AI aid, highlighting the importance of developing transparent AI systems within healthcare.

Keywords: artificial intelligence; trust in automation; human-computer interaction; uncertainty communication; visualization; pharmacist; medication errors; patient safety;

INTRODUCTION

Pharmacists are responsible for dispensing medication products that are consistent with the prescription ordered by healthcare prescribers. To do this, they perform a verification task where they compare the filled medication inside the vial to the product on the prescription label. Due to technology limitations and the high workload faced by pharmacy staff, dispensing errors can occur [1,2]. Dispensing errors, defined as instances when patients receive an incorrect drug or dose, contribute significantly to preventable adverse drug events that result in 700,000 emergency room visits and 100,000 hospital admissions each year [3]. To improve patient health outcomes, minimize unnecessary healthcare costs, and alleviate pharmacists' overload, it is essential to provide pharmacists with reliable tools to minimize the risk of incorrect dispensing [4].

Since the 1990s, the implementation of barcode scanning systems has been advocated as a means to reduce medication errors [5]. The adoption of such systems within pharmacies and broader healthcare environments has led to a notable reduction in medication errors [6-9]. Nevertheless, research indicates that barcode scanning systems are not immune to workarounds and human errors [10-12]. Moreover, these systems do not adequately address the challenges faced by overburdened pharmacists [13-16].

In response to these challenges, pill counting and verification/recognition systems using image classification technologies have emerged [17-20]. Innovations like Eyecon and VIVID employ vision-based methods to count medications placed on the tray. More recently, advancements have been made in automated pill identification through feature engineering. Yu, Chen, and Kamata [21] and Yu et al. [22] proposed an automatic pill recognition method based on pill imprints, achieving an accuracy of 86.01% and 90.46%, respectively. Caban et al. [17] used a modified shape distribution technique to determine the shape, color, and imprint of a pill to identify the drug. The proposed technique was evaluated with 568 of the most prescribed drugs in the United States and achieved a 91.13% accuracy.

The advent of deep learning has further enhanced the capabilities of automated pill recognition systems [4, 23]. For instance, Larios Delgado et al. [4] developed a pill recognition method using two deep-learning models. They used a deep Convolutional Neural Network (CNN) model for pill blob detection to isolate the pill from the background and then passed the output to a deep-learning-based classifier to identify the five most likely pills with 94% accuracy [4]. Similarly, Wong et al. [24] proposed a Deep Convolutional Network model and achieved a mean accuracy of 95.35%. Lester et al. [23] trained a ResNet-18 deep-learning model to predict the labeled features of a medication product using an image showing medication inside a filled

prescription vial. In a test set containing 65,274 images of 345 unique oral drug products, the overall macro-average precision, i.e., the mean of precision values for each class, was 98.5%.

Despite the impressive strides in model accuracy, realizing the potential of these technologies is only possible if people establish appropriate trust in them. Trust in automation, defined as "the attitude that an (autonomous) agent will help achieve an individual's goals in situations characterized by uncertainty and vulnerability [25]" is one of the most crucial factors determining the use of automation [26, 27]. There is a growing body of research examining people's trust in autonomous and robotic technologies in various domains, including transportation [28, 29], healthcare [30], education [31], and defense [31]. This study, therefore, aimed to investigate pharmacists' trust in automated pill verification technology, which is designed to assist with the critical task of medication dispensing.

METHODS

This research was exempt from IRB oversight by the University of Michigan Institutional Review Board, and all participants signed a prospective agreement to participate.

Recruitment and Participants

Recruitment emails were dispatched to pharmacists through the Minnesota Pharmacy Practice-Based Research Network and the University of Michigan College of Pharmacy Pharmacist Preceptor Network. To meet the inclusion criteria, pharmacists were required to 1) be licensed pharmacists in the United States, 2) be at least 18 years old, and 3) have access to a laptop or desktop computer with a webcam. Pharmacists who (1) require assistive technology to use the computer, (2) wear eyeglasses with more than one power, (3) have uncorrected cataracts, intraocular implants, glaucoma, or permanently dilated pupils, and/or (4) have eye movement or alignment abnormalities (e.g., lazy eye, strabismus, nystagmus) were excluded from participation in the study. A total number of N = 30 licensed pharmacists in the United States completed the study (see *Figure 1*).

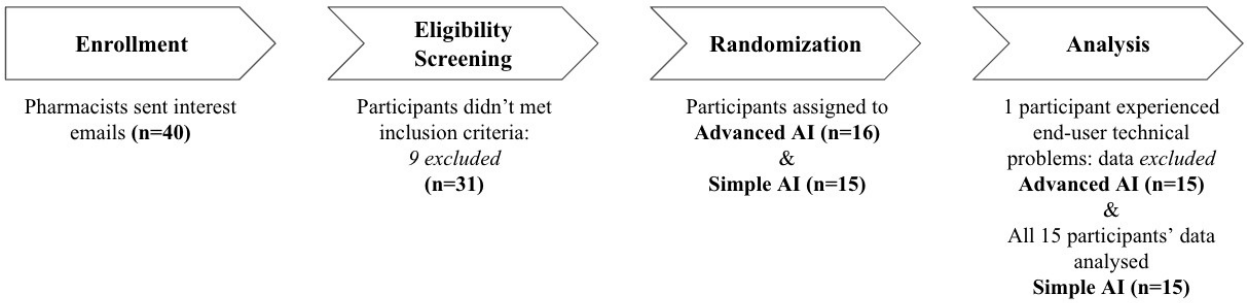


Figure 1. Participant Recruitment Timeline.

AI Model

The AI model used in this study is a Bayesian neural network (BNN) that predicts the National Drug Code (NDC), a unique identifier assigned by the Food and Drug Administration (FDA) to catalog drug products in the United States [33, 34]. The BNN used a technique known as random dropout [35], applied to a ResNet-34 Convolutional Neural Network (CNN) architecture [36] to estimate the probability associated with each NDC (i.e., each class). The model produced 50 different predictions, where each prediction is a probability vector that quantifies the probabilities that an input belongs to each of the NDCs. The predicted NDC was then attained by finding the highest average probability derived from the fifty predictions.

To train the AI model, we acquired a dataset of 432,974 images from a mail-order pharmacy in the United States. This pharmacy uses a robotic system that counts pills, fills and labels the vial, captures the images of the contents, and seals the vial with a cap. The image dataset consists of one year's worth of robot-captured images of oral medications, such as tablets and capsules, inside prescription vials filled by the robotic system. Each image in the dataset is associated with an NDC label and various attributes of color, shape, size, manufacturer, tablet scoring, and imprint. The number of images available for each NDC varied, ranging from 3 to 12,105, with a median of 540. The medications featured in these datasets were classified into twelve different colors and seven distinct shapes. The detailed classification is shown in *Table 1*.

TABLE 1: Percentage of medication characteristics (colors on the left, shapes on the right) featured in the training dataset.

Colors	Percentage (%)	Shapes	Percentage (%)
White	42.1	Round	49.6
Yellow	12.3	Oval	33.4
Pink	9.1	Capsule	16.2

Orange	7.1	Hexagon-6-sided	0.4
Multi-color	5.9	Triangle	0.3
Green	5.2	Trapezoid	0.1
Red	5.2	Pentagon-5-sided	0
Blue	4.8		
Brown	3.8		
Purple	3.1		
Turquoise	0.7		
Gray	0.7		

Experimental Testbed and Stimuli

In the experiment, participants performed a pill verification task with the help of an imperfect AI aid that recommends whether to accept or reject a filled medication. The participant's task was to verify whether the filled medication matched the reference image. If the reference image and filled medication did not match, the correct action was to click "reject." If the reference image and filled medication matched, the correct action was to click "accept."

The user interface was designed following pharmacists' feedback from a focus group study conducted by the research team [34]. The interface displayed an image of a filled medication, a reference image, prescription information, and AI aids. There were two types of AI aids powered by the same AI model: one aid augmented with an uncertainty plot indicating the degree of (un)certainty of the AI recommendation, and the other aid without this feature. Both AI aids recommend the action the pharmacist should take, using four checkboxes. A recommendation to accept was indicated when all four checkboxes turned green (see *Figure 2*); otherwise, the recommendation was to reject. For the AI aid with the uncertainty plot, an additional histogram was integrated (*Figure 3*). The histogram displayed the distribution of the fifty probabilities for the predicted NDC, generated by the fifty predictions. The purpose of the histogram was to provide a visual representation of the certainty or uncertainty level associated with the AI's NDC prediction.

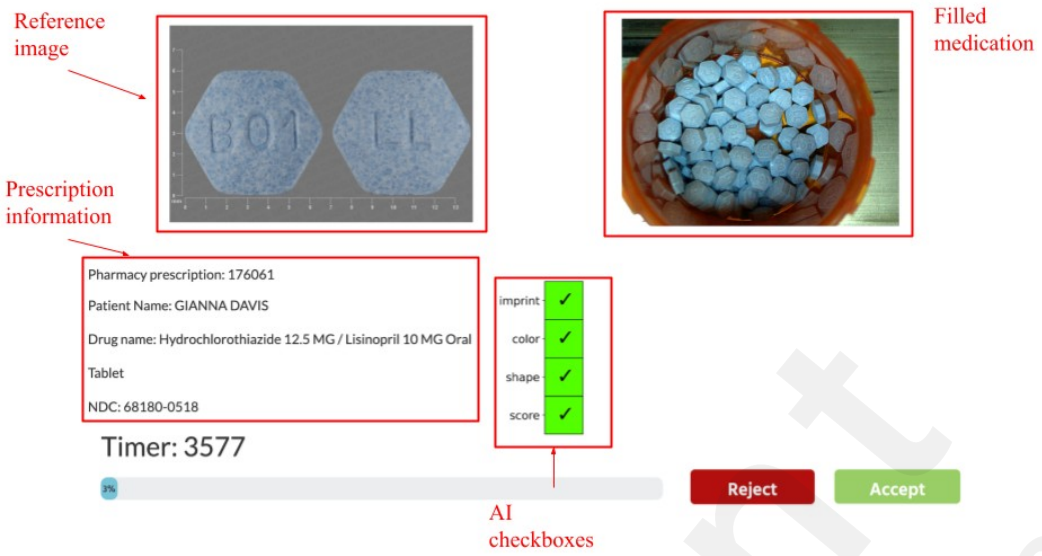


Figure 2. Interface for the AI aid without the uncertainty plot. Checkboxes indicate AI's recommendation. When all four checkboxes are green, the AI advises to accept otherwise, to reject.

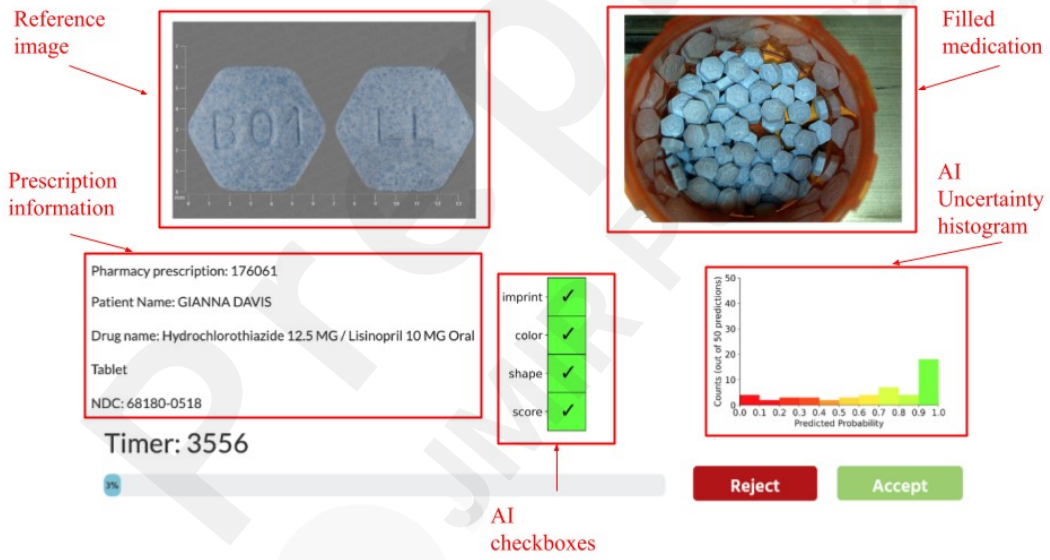


Figure 3. Interface for the AI aid with the uncertainty plot. In addition to *Figure 2*, the histogram shows the distribution of the 50 predicted probabilities associated with the predicted NDC.

With the help of an AI aid, participants performed a block of 100 pill verification trials. The experimental stimuli for the 100 trials, including the reference NDC and the filled medication, were carefully selected from the dataset of 432,974 images. The selection process ensured a broad representation of colors and shapes, while blurry images were excluded to maintain clarity. To minimize learning effects, each reference NDC was intentionally shown no more than twice throughout the experiment.

Furthermore, the AI aid was not perfect for the 100 trials, i.e., it offered incorrect

recommendations occasionally. Based on the Signal Detection Theory, we mapped out the relationship between the AI's recommendation and the true state of the world [37]. In the context of this experiment, a signal in the world was an incorrectly filled medication. The extent to which the AI recommended rejecting an incorrectly filled medication reflects its ability to detect the signal. The combination of the state of the world and the AI's recommendation resulted in four potential outcomes: AI rejects the incorrect drug (hit), AI approves the incorrect drug (miss), AI rejects the correct drug (false alarms), and AI approves the correct drug (correct rejections), as shown in *Figure 4*.

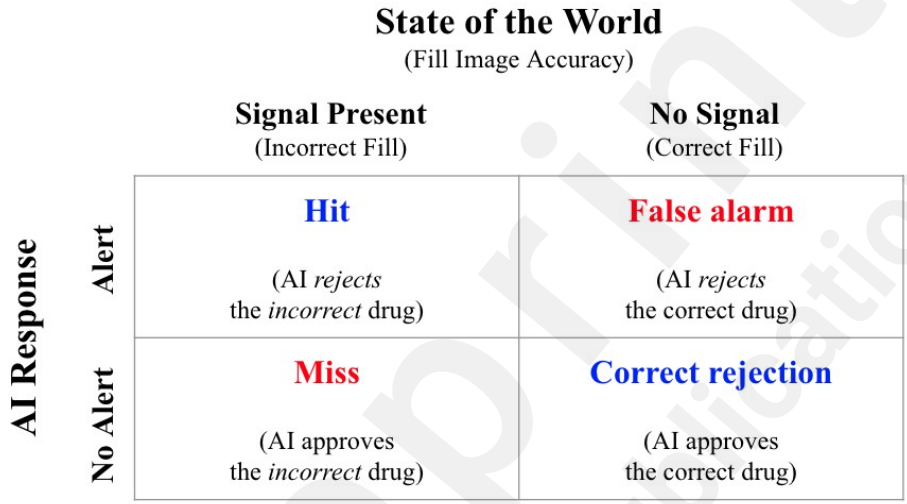


Figure 4. Four potential AI performance patterns according to Signal Detection Theory.

Benchmarking prior literature [38], the base rate was set to be 24%, i.e., 24% of the trials contained wrongly filled medication. The AI accuracy was set as 82% to ensure that AI would be perceived as useful while providing sufficient misses and false alarms [39]. By combining the filled image accuracy and AI accuracy, there were 60 cases of AI approving the correct drug, 22 cases of AI rejecting the incorrect drug, 2 cases of AI approving the incorrect drug, and 16 cases of AI rejecting the correct drug.

After each trial, participants reported their trust in the recognition AI on a visual analog scale with the leftmost point labeled "Not at all trust" and the rightmost point labeled "Completely trust" [40-42] (*Figure 5*).

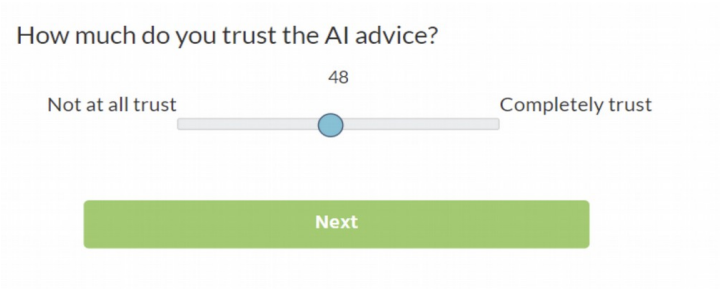


Figure 5. Participants reported their trust in the recognition AI on a visual analog scale.

Experimental Design

The experiment employed a mixed design. The between-subjects factor was the type of AI aid, distinguished by the presence or absence of an uncertainty plot. The within-subjects factor was the four potential outcomes: AI rejects the incorrect drug, AI rejects the correct drug, AI approves the incorrect drug, and AI approves the correct drug (*Figure 4*).

Half of the *N*=30 participants used the AI aid without the uncertainty plot, and the other half used the AI aid with the uncertainty plot. Each participant completed two blocks of 100 trials each. One block involved using the AI aid (either with or without the uncertainty plot), and the other block required participants to perform the task manually. The order of the two blocks was counterbalanced. Additionally, benchmarking prior literature [43], the trial sequence was fixed for the 100 trials in each block.

As this study is focused on the pharmacists' trust in AI, data from the manual task block were excluded from the analysis, concentrating the study's findings on interactions involving the AI aid.

Measures

Trust propensity. Before the experiment, we measured participants' trust propensity using the six-item survey used by Merritt et al. [44]. Trust propensity is "a stable, trait-like tendency to trust or not trust others [45]", and the propensity to trust machines reflects a person's tendency to trust machines in general rather than in a particular machine.

End Trust. End trust, *Trust (100)*, is the participant's final trust rating after interacting with the AI help scenario.

Average Trust. Average trust denotes the mean of moment-to-moment trust ratings collected throughout the experiment.

$$Average\ Trust = \sum_{i=2}^{100} Trust(i)$$

Trust change. After each trial *i*, participants report their *trust*(*i*) in the AI. We calculate a trust change as:

$$Trust\ change(i) = Trust(i) - Trust(i-1), where\ i = 2,3,...100$$

Since the moment-to-moment trust is reported after each trial, 99 trust changes are obtained from each participant.

Experimental procedure

Before each experiment, participants had a brief meeting with a member of the study team to ensure that the physical environment, including lighting conditions, was suitable for the experiment. Subsequently, the pharmacists were directed to Labvanced's website and presented with a video tutorial that explained how to perform a simulated medication verification task using the testbed interface. Pharmacists were informed that the objective of the task was to determine whether an image of a filled medication bottle matched a known reference image. The tutorial also explained the two AI aids.

Before engaging in the verification task, participants were directed to complete a demographics survey and a trust propensity survey [44]. Additionally, they went through a set of calibration procedures for the
eye-tracking software. Participants then completed a block of 100 verification trials using an AI aid – either with or without the uncertainty plot – and another block of 100 verification trials manually, conducted in a counterbalanced order. Upon completion of the 200 trials, participants filled out a post-experimental survey and answered non-mandatory free-response feedback questions. All participants finished each block within the time limit of 60 minutes.

After completing all the tasks and surveys described above, participants were invited to a 30-minute debriefing session with one of the study team members. Study team members described 6 concepts of automation evaluation: observability, predictability, directing attention, exploring solution space, adaptability, and calibrated trust [46], and provided an example scenario for each concept. After each description and example, the participants were asked to provide their thoughts on how the concept relates to our system.

Statistical Analysis

Participants' trust propensity, end trust, and trust change when using both types of AI aids were analyzed. First, we conducted a descriptive analysis of the participants' trust propensity. Then, we compared their end trust towards the AI aid with or without uncertainty plot using an

independent samples *t*-test. Finally, we analyzed how trust increased and decreased after participants experienced each of the four AI performance patterns using mixed-linear models with random intercept. Regression *t*-tests were conducted to compare the means between the two AI aids for each AI performance pattern. The Kenward-Roger method was used for estimating degrees of freedom. Mixed model GLM tests were conducted using the R lme4 package [47]. All statistical analyses were carried out using R statistical software, version 4.2.2 [48].

RESULTS

Trust Propensity

Participants had an average trust propensity score of 72 out of 100 (SD = 18.08), indicating the participants generally had a positive attitude towards trusting automated technologies. There was no significant difference between the two AI aids ($t(28) = -0.854$, $P = .2$).

End-Trust Towards AI Aid

Participants trusted the AI aid with the uncertainty plot significantly more than the AI aid without the plot at the end of the experiment (*t(28) = -1.854, P = .037*) (*Figure 6*).

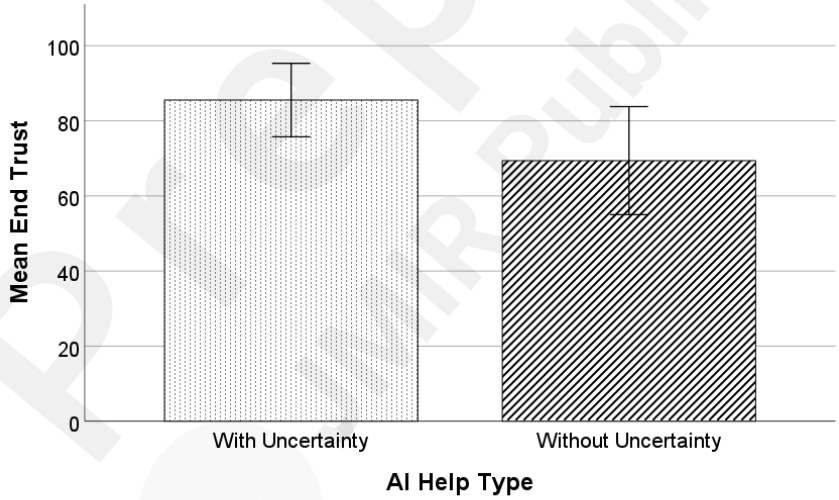


Figure 6. Mean end trust by AI aid with or without the uncertainty plot. The error bars represent +/- 2 standard errors.

Average Trust Towards AI Aid

There seemed to be a trend that the average trust was greater for the AI aid with uncertainty plot compared to the aid without the plot (see *Figure 7*). However, the difference did not reach statistical significance (*t(28) = -1.036, P = .155*).

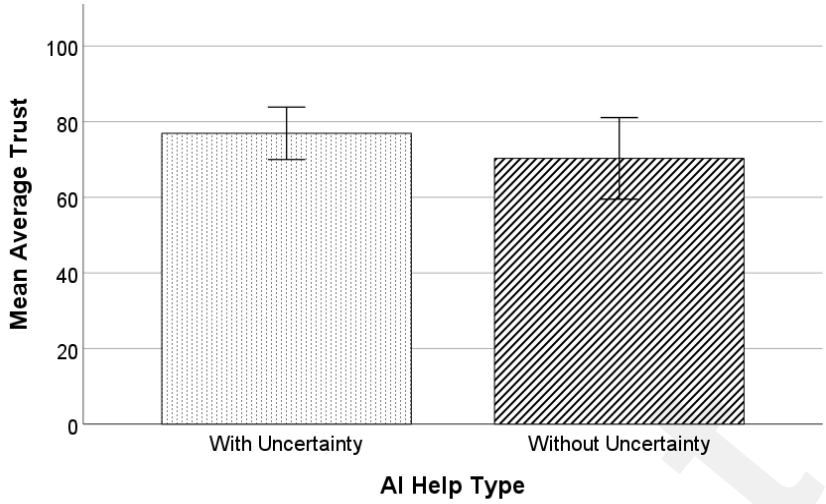


Figure 7. Mean average trust by AI aid with or without the uncertainty plot. The error bars represent +/- 2 standard errors.

Trust Change

Figure 8 shows the trust change after participants experienced the four AI performance patterns. When AI approved the correct drug, there was a significantly greater trust increment when participants used the AI aid with the uncertainty plot compared to the one without (*t(78.98) = 3.93, P < .001*); When AI approved the incorrect drug, there was a significantly greater trust decrement using the AI aid with the uncertainty plot (*t(2939.72) = -4.78, P < .001*). Interestingly, when AI rejected the incorrect drug, we observed a decrement of trust for participants using the AI aid with the uncertainty plot (*t(509.77) = -3.96, P < .001*). Overall,

participants using the AI aid with the uncertainty plot displayed a large magnitude of trust adjustment. In addition, we observed a significant “negativity bias” in that the magnitude of trust change when AI made an error (i.e., AI approves the incorrect drug or AI rejects the correct drug) was significantly larger than the magnitude of trust adjustment when AI provided correct recommendations (generalized linear model test, *z = -11.30, P < .001*).

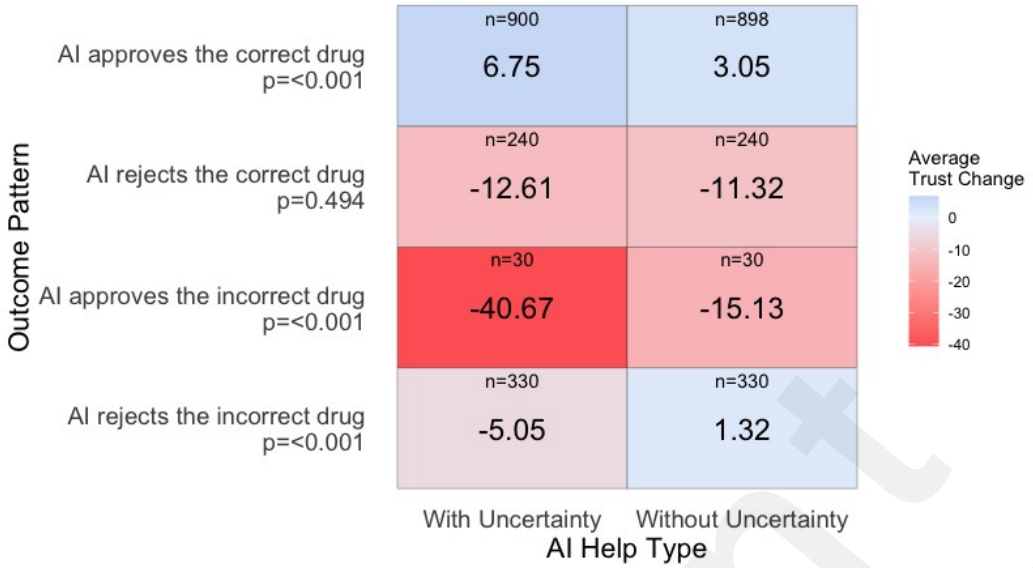


Figure 8. Trust change for different AI performance patterns. Red shades represent negative trust change, and blue shades represent positive trust change.

As we measured participants' trust towards AI continuously, we calculated the autocorrelation between the trust ratings. Autocorrelation measures the relationship between a variable's current value and its past values in a time series. *Figure 9* shows the mean autocorrelation as a function of time separation between the ratings. For both AI aids, the correlation decreased as the time separation increased. The AI aid with the uncertainty plot had a lower mean autocorrelation compared to the aid without the uncertainty plot (*Figure 9*).

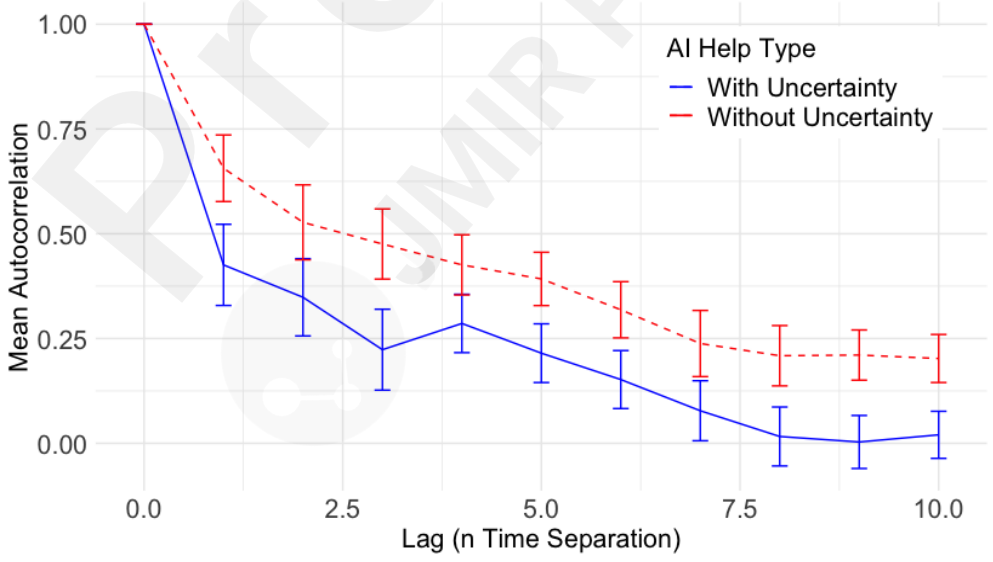


Figure 9. Autocorrelation of trust as a function of time separation. The blue solid line represents AI aid with the uncertainty plot, and the red dashed line represents the AI aid without the uncertainty plot. The error bars represent +/- 2 standard errors.

DISCUSSION

The propensity to trust automation refers to an individual's general inclination to trust automated/autonomous systems shaped by their past experiences and future expectations [44, 45]. Research has documented varying levels of trust propensity among individuals. In an early study employing 69 college students (average age=25), Merritt et al. [44] found an average trust propensity of 3.56 (SD = 0.6) on a 7-point Likert scale. More recently, Montag et al. [49] surveyed 289 participants aged between 18 and 70 and reported their propensity to trust automation to be 4.98 (SD =1.06) after converting to a 7-point Likert scale. Miller et al. [50] reported a similar trust propensity score of 4.97 (SD = 1.21) from a smaller cohort of 28 participants aged between 18 to 60. Another investigation by Yang, Guo, and Schemanske [51] with 75 adults (mean age = 23.4) divided into three groups reported trust propensity scores of 72.6 (SD = 14.8), 69.4 (SD = 10.4), and 69.4 (SD = 14.4), equivalent to average scores of 5.08, 4.86 and 4.86 on a 7-point Likert scale.

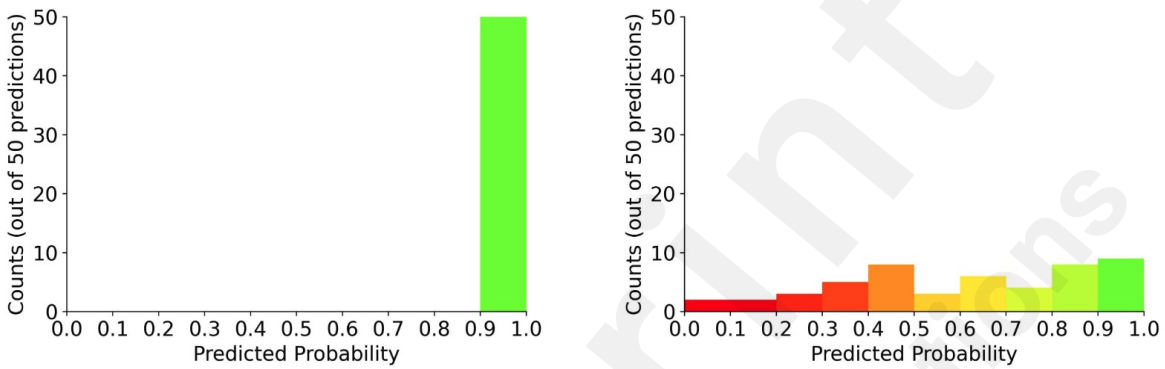
Our finding on trust propensity aligns with recent findings [49-51], indicating that pharmacists generally have a favorable disposition towards trusting automation, with an average rating of 72 on a 100-point scale, or 5.03 on a 7-point Likert Scale. This positive attitude may be attributed to the frequent use of automated technologies, such as barcode scanners and pill counters in pharmacies [5, 6]. Additionally, as expected, we did not reveal a significant difference between the groups using different AI aids because the participants were randomly assigned to use either of the AI aids.

Regarding pharmacists' end trust, our findings reveal that the AI aid with the uncertainty plot significantly enhanced the end trust scores. We attribute this enhancement to the increased transparency achieved through the presentation of a histogram showing the distribution of 50 predicted probabilities. While AI advancements promise to improve human performance, a prevailing issue is the perception of AI as a "black box." This lack of transparency contributes to a lack of trust in AI and can undermine team performance [52-54]. The higher end-trust observed in participants using the AI aid with the uncertainty plot indicates that making the AI more transparent by revealing its decision-making process can foster a lighter level of trust in automation.

With respect to the dynamics of trust, i.e., moment-to-moment trust change, when AI approved the correctly filled bottle, we noted trust increments for both AI aids. Furthermore, the inclusion of uncertainty information led to a larger increment in trust compared to when such information was absent. When AI mistakenly approved the incorrect drug, we observed a significant trust decrement for both AI aids, potentially attributed to the adverse outcome associated with the wrong medication. Furthermore, the trust decrement was significantly larger when the uncertainty information was shown. This pronounced trust decrement could have resulted directly from the

distribution of the histogram: participants were shown a histogram indicating a high level of certainty (illustrated in *Figure 10a*). Therefore, participants may have perceived the error made by the AI aid as a “confident” error and therefore reduced their trust even more. Studies examining likelihood alarms reported that

highly likely alarms (i.e., “confident” alarms) engender a greater decline in momentary trust upon automation failures [41, 55].



a b

Figure 10. Uncertainty plot with narrower (left) and wider (right) interquartile range. Interquartile range is a measure of statistical dispersion, or how spread out the data points are.

Intriguingly, when AI rejected an incorrectly filled bottle, the absence of uncertainty information resulted in an increase in trust, whereas the presence of such information led to a decrease in trust. Such contrasting results could have stemmed from the uncertainty plots influencing the participants’ decision-making process. When AI approved the correct drug, all uncertainty plots presented to participants showed a consistent solid green bar (*Figure 10a*). However, when AI rejected the incorrect drug, the interquartile range of the uncertainty plot was broader, indicating the lack of certainty (see *Figure 11*). 16 out of 22 uncertainty plots displayed a wider spread with mixed color bars (illustrated in *Figure 10b*). This ambiguity might unintentionally cause the human participants to doubt the capability of the AI aid with the uncertainty plot, resulting in a decrease in their trust.

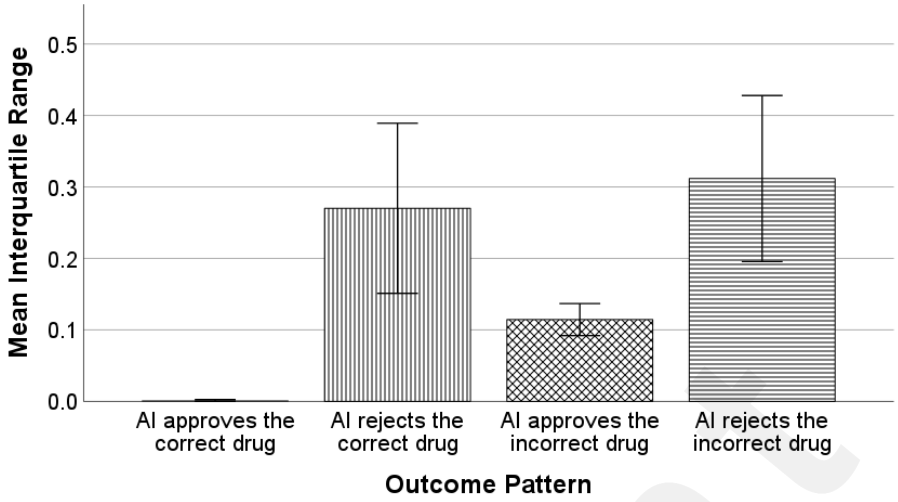


Figure 11. Mean interquartile range by outcome pattern. The error bars represent +/- 2 standard errors.

Finally, when AI rejected the correctly filled bottle, trust decrements occurred, with no significant differences between the two AI aids. If this circumstance happened in the real world, the pharmacists would re-inspect the filled prescription. It will likely lead to an increased workload, fatigue, and stress, which could potentially lead to a lower quality of work and a higher frequency of errors [56, 57].

The observed trends in the moment-to-moment dynamics of trust indicate a greater degree of trust adjustment when participants were assisted by the AI with the uncertainty plot. This observation is further confirmed by the autocorrelation analysis. Specifically, the trust autocorrelation plot (*Figure 9*) reveals a lower autocorrelation between trust ratings when the uncertainty information was presented. This suggests that current trust levels were less influenced by past trust levels, implying more significant changes in trust from moment to moment. Pharmacists relied less on previous trials and more on the information presented in the present trial, highlighting the advantages of a more transparent display.

In addition, for both AI aids, participants displayed a larger trust decrement due to incorrect automation predictions. Even though these observations may seem alarming initially, they align with the prior literature addressing negativity bias. The study suggests that failure in automation typically has a more significant negative impact on trust than a positive impact from successful automation [40, 51].

Conclusion

Dispensing errors are significant contributors to adverse drug events, which lead to considerable healthcare expenses and harm to patients. Despite progress made in developing automated technologies to aid pill verification, pharmacists' trust in these systems has not been thoroughly investigated. Our research represents an initial exploration into pharmacists' trust in automated pill verification technology, marking a significant step in understanding the integration of such systems into healthcare settings.

Our findings reveal that pharmacists have a favorable disposition towards trusting automation, likely attributed to their frequent use of automated technologies in their work. Moreover, providing uncertainty information about the AI's recommendation significantly boosts pharmacists' trust in the AI aid, highlighting the importance of transparency in AI development. The dynamics of trust vary depending on the AI's performance. Pharmacists using the AI aid with the uncertainty plot had a significantly larger trust increment when AI approved the correct drug and a significantly larger trust decrement when AI approved the incorrect drug. Intriguingly, the absence of the uncertainty plot led to an increase in trust when AI correctly rejected an incorrect drug, whereas the presence of the plot resulted in a decrease in trust under the same circumstances. In addition, a pronounced "negativity bias" was observed, where the degree of trust reduction when AI made an error exceeded the trust gain when AI made a correct decision.

The study should be viewed in light of the following limitations: First, being pioneering research in this area, we did not strictly control the interquartile range of the uncertainty plot. Future investigation is needed to examine the effects of presenting different distributions in the uncertainty plot on pharmacists' trust. Second, the uncertainty plot used in the present study only showed the distribution of the fifty probabilities for the predicted NDC. Further research could improve the uncertainty plot presentation, ensuring easier interpretation of the AI's recommendation.

Acknowledgments

This research study was supported by the National Library of Medicine of the National Institutes of Health under award number R01LM013624. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

REFERENCES

1. Chui MA, Look KA, Mott DA. The association of subjective workload dimensions on quality of care and pharmacist quality of work life. Research in Social and Administrative Pharmacy. 2014;10(2):328-340.
2. Chui MA, Mott DA. Community pharmacists' subjective workload and perceived task performance: a human factors approach. Journal of the American Pharmacists Association. 2012;52(6):e153-e160.
3. Reiner G, Pierce SL, Flynn J. Wrong drug and wrong dose dispensing errors identified in pharmacist professional liability claims. Journal of the American Pharmacists Association. 2020;60(5):e50-e56.
4. Larios Delgado N, Usuyama N, Hall AK, et al. Fast and accurate medication identification. NPJ digital medicine. 2019;2(1):10.
5. Bates DW. Using information technology to reduce rates of medication errors in hospitals. Bmj. 2000;320(7237):788-791.
6. Poon EG, Cina JL, Churchill W, et al. Medication dispensing errors and potential adverse drug events before and after implementing bar code technology in the pharmacy. Annals of internal medicine. 2006;145(6):426-434.
7. Poon EG, Keohane CA, Yoon CS, et al. Effect of bar-code technology on the safety of medication administration. New England Journal of Medicine. 2010;362(18):1698-1707.
8. Paoletti RD, Suess TM, Lesko MG, et al. Using bar-code technology and medication observation methodology for safer medication administration. American journal of health-system pharmacy. 2007;64(5):536-543.
9. Young D. Veterans Affairs bar-code-scanning system reduces medication errors. American journal of health-system pharmacy: AJHP: official journal of the American Society of Health-System Pharmacists. 2002;59(7):591-592.
10. Debono DS, Greenfield D, Travaglia JF, et al. Nurses' workarounds in acute healthcare settings: a scoping review. BMC health services research. 2013;13:1-16.
11. Koppel R, Wetterneck T, Telles JL, Karsh B-T. Workarounds to barcode medication administration systems: their occurrences, causes, and threats to patient safety. Journal of the American Medical Informatics Association. 2008;15(4):408-423.
12. Oldland AR, Golightly LK, May SK, Barber GR, Stolpman NM. Electronic inventory systems and barcode technology: impact on pharmacy technical accuracy and error liability. Hospital pharmacy. 2015;50(1):034-041.

13. Buchanan TL, Barker KN, Gibson JT, Jiang BC, Pearson RE. Illumination and errors in dispensing. American journal of hospital pharmacy. 1991;48(10):2137-2145.
14. Gidman W. Increasing community pharmacy workloads in England: causes and consequences. International journal of clinical pharmacy. 2011;33:512-520.
15. Lea VM, Corlett SA, Rodgers RM. Workload and its impact on community pharmacists' job satisfaction and stress: a review of the literature. International Journal of Pharmacy Practice. 2012;20(4):259-271.
16. Peterson G, Wu M, Bergin J. Pharmacists' attitudes towards dispensing errors: their causes and prevention. Journal of clinical pharmacy and therapeutics. 1999;24(1):57-71.
17. Caban JJ, Rosebrock A, Yoo TS. Automatic identification of prescription drugs using shape distribution models. IEEE; 2012:1005-1008.
18. Hamilton RN. Pharmacy pill counting vision system. Google Patents; 2004.
19. Popovich Jr J, Jordan ML. Automated drug discrimination during dispensing. Google Patents; 2011.
20. Wootton JR, Reznack VV, Hobson G. Pharmaceutical pill recognition and verification system. Google Patents; 2003.
21. Yu J, Chen Z, Kamata S-i. Pill recognition using imprint information by two-step sampling distance sets. IEEE; 2014:3156-3161.
22. Yu J, Chen Z, Kamata Si, Yang J. Accurate system for automatic pill recognition using imprint information. IET Image Processing. 2015;9(12):1039-1047.
23. Lester CA, Li J, Ding Y, Rowell B, Yang JX, Kontar RA. Performance evaluation of a prescription medication image classification model: an observational cohort. NPJ Digital Medicine. 2021;4(1):118.
24. Wong YF, Ng HT, Leung KY, Chan KY, Chan SY, Loy CC. Development of fine-grained pill identification algorithm using deep convolutional network. Journal of biomedical informatics. 2017;74:130-136.
25. Lee JD, See KA. Trust in automation: Designing for appropriate reliance. Human factors. 2004;46(1):50-80.
26. Hoff KA, Bashir M. Trust in automation: Integrating empirical evidence on factors that influence trust. Human factors. 2015;57(3):407-434.
27. Schaefer KE, Chen JY, Szalma JL, Hancock PA. A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. Human factors. 2016;58(3):377-400.

28. Azevedo-Sa H, Jayaraman SK, Esterwood CT, Yang XJ, Robert Jr LP, Tilbury DM. Real-time estimation of drivers' trust in automated driving systems. International Journal of Social Robotics. 2021;13(8):1911-1927.
29. Lee JD, Kolodge K. Exploring trust in self-driving vehicles through text analysis. Human factors. 2020;62(2):260-277.
30. Gombolay M, Yang XJ, Hayes B, et al. Robotic assistance in the coordination of patient care. The International Journal of Robotics Research. 2018;37(10):1300-1316.
31. Nazaretsky T, Ariely M, Cukurova M, Alexandron G. Teachers' trust in AI-powered educational technology and a professional development program to improve it. British journal of educational technology. 2022;53(4):914-931.
32. Guo Y, Yang XJ, Shi C. Enabling Team of Teams: A Trust Inference and Propagation (TIP) Model in Multi-Human Multi-Robot Teams. Proceedings of Robotics: Science and Systems (RSS '23). 2023;
33. Chen Q, Kontar RA, Nouiehed M, Yang J, Lester C. Rethinking Cost-sensitive Classification in Deep Learning via Adversarial Data Augmentation. arXiv preprint arXiv:220811739. 2022;
34. Zheng Y, Rowell B, Chen Q, et al. Designing Human-Centered AI to Prevent Medication Dispensing Errors: Focus Group Study With Pharmacists. JMIR Formative Research. 2023;7(1):e51921.
35. Gal Y, Ghahramani Z. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. PMLR; 2016:1050-1059.
36. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. 2016:770-778.
37. Tanner Jr WP, Swets JA. A decision-making theory of visual detection. Psychological review. 1954;61(6):401.
38. Reilley S, Grasha AF, Matthews G, Schafer J. Automatic-controlled information processing and error detection in a simulated pharmacy-verification task. Perceptual and motor skills. 2003;97(1):151-174.
39. Wickens CD, Dixon SR. The benefits of imperfect diagnostic automation: A synthesis of the literature. Theoretical Issues in Ergonomics Science. 2007;8(3):201-212.
40. Manzey D, Reichenbach J, Onnasch L. Human performance consequences of automated decision aids: The impact of degree of automation and system experience. Journal of Cognitive Engineering and Decision Making. 2012;6(1):57-87.
41. Yang XJ, Unhelkar VV, Li K, Shah JA. Evaluating effects of user experience and

system transparency on trust in automation. 2017:408-416.

42. Yang XJ, Schemanske C, Searle C. Toward quantifying trust dynamics: How people adjust their trust after moment-to-moment interaction with automation. Human Factors. 2023;65(5):862-878.
43. Lee J, Moray N. Trust, control strategies and allocation of function in human-machine systems. Ergonomics. 1992;35(10):1243-1270.
44. Merritt SM, Heimbaugh H, LaChapell J, Lee D. I trust it, but I don't know why: Effects of implicit attitudes toward automation on trust in an automated system. Human factors. 2013;55(3):520-534.
45. Merritt SM, Ilgen DR. Not all trust is created equal: Dispositional and history-based trust in human-automation interactions. Human factors. 2008;50(2):194-210.
46. McDermott P, Dominguez C, Kasdaglis N, Ryan M, Trhan I, Nelson A. Human-machine teaming systems engineering guide. MITRE Corp Bedford MA Bedford United States. 2018;
47. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:14065823. 2014;
48. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2019;
49. Montag C, Kraus J, Baumann M, Rozgonjuk D. The propensity to trust in (automated) technology mediates the links between technology self-efficacy and fear and acceptance of artificial intelligence. Computers in Human Behavior Reports. 2023;11:100315.
50. Miller L, Kraus J, Babel F, Baumann M. More than a feeling—interrelation of trust layers in human-robot interaction and the role of user dispositions and state anxiety. Frontiers in psychology. 2021;12:592711.
51. Yang XJ, Guo Y, Schemanske C. From trust to trust dynamics: Combining empirical and computational approaches to model and predict trust dynamics in human-autonomy interaction. Human-Automation Interaction: Transportation. Springer; 2022:253-265.
52. De Visser EJ, Pak R, Shaw TH. From 'automation' to 'autonomy': the importance of trust repair in human–machine interaction. Ergonomics. 2018;61(10):1409-1427.
53. Lyons JB, Sadler GG, Koltai K, et al. Shaping trust through transparent design: theoretical and experimental guidelines. Springer; 2017:127-136.
54. Luo R, Du N, Yang XJ. Evaluating effects of enhanced autonomy transparency on

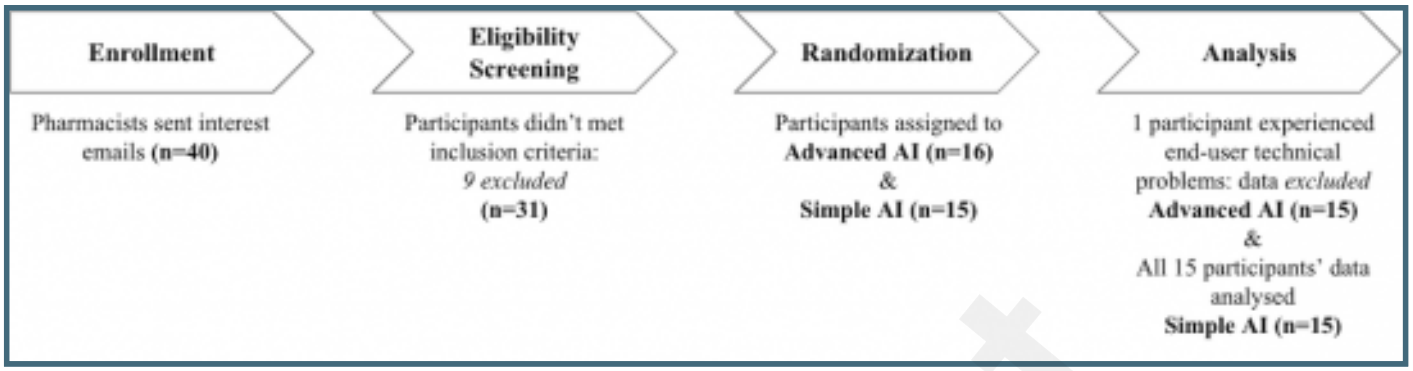
trust, dependence, and human-autonomy team performance over time. International Journal of Human–Computer Interaction. 2022;38(18-20):1962-1971.

55. Sorkin RD, Kantowitz BH, Kantowitz SC. Likelihood alarm displays. Human Factors. 1988;30(4):445-459.
56. Carayon P, Gurses AP. Nursing workload and patient safety—a human factors engineering perspective. Patient safety and quality: An evidence-based handbook for nurses. 2008;
57. Watterson TL, Look KA, Steege L, Chui MA. Operationalizing occupational fatigue in pharmacists: An exploratory factor analysis. Research in Social and Administrative Pharmacy. 2021;17(7):1282-1287.

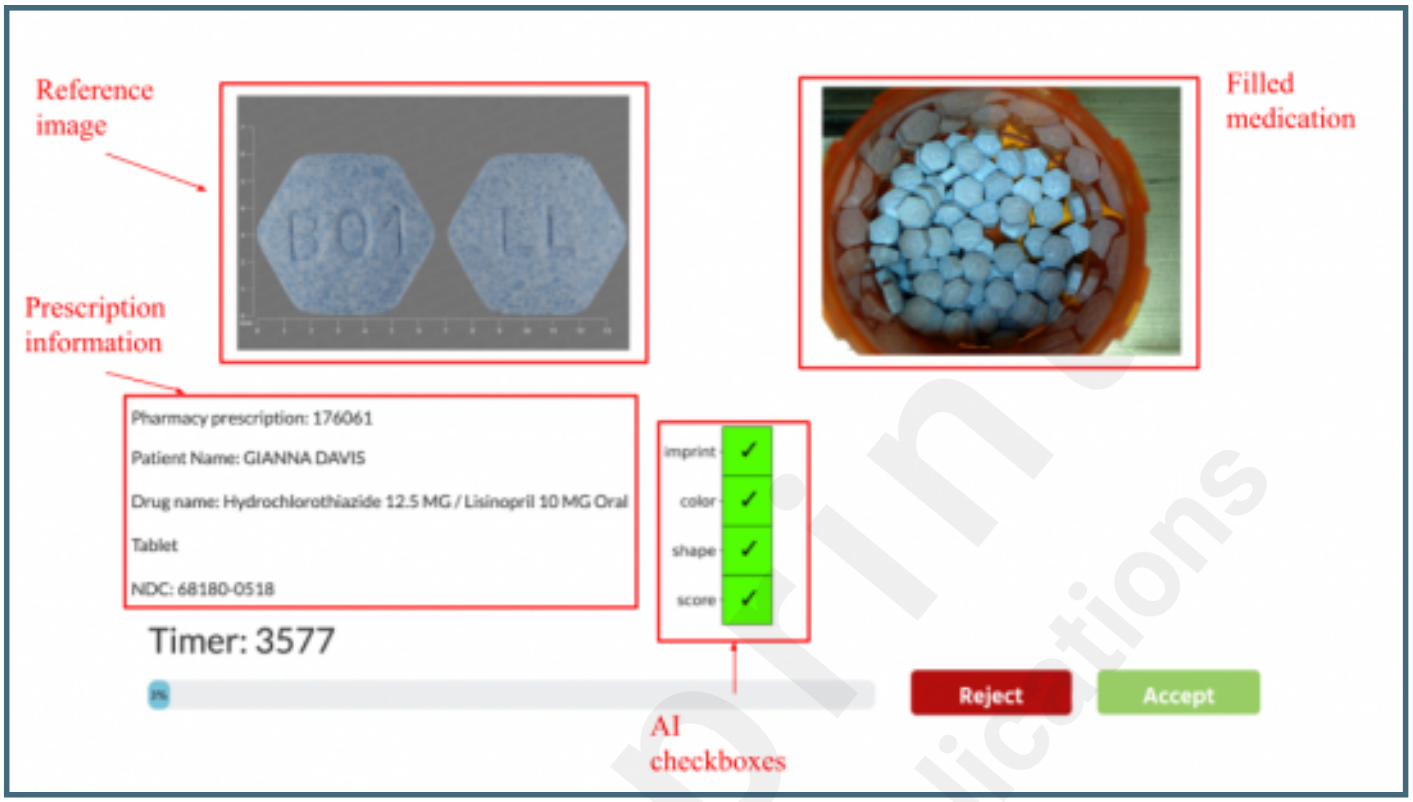
Supplementary Files

Figures

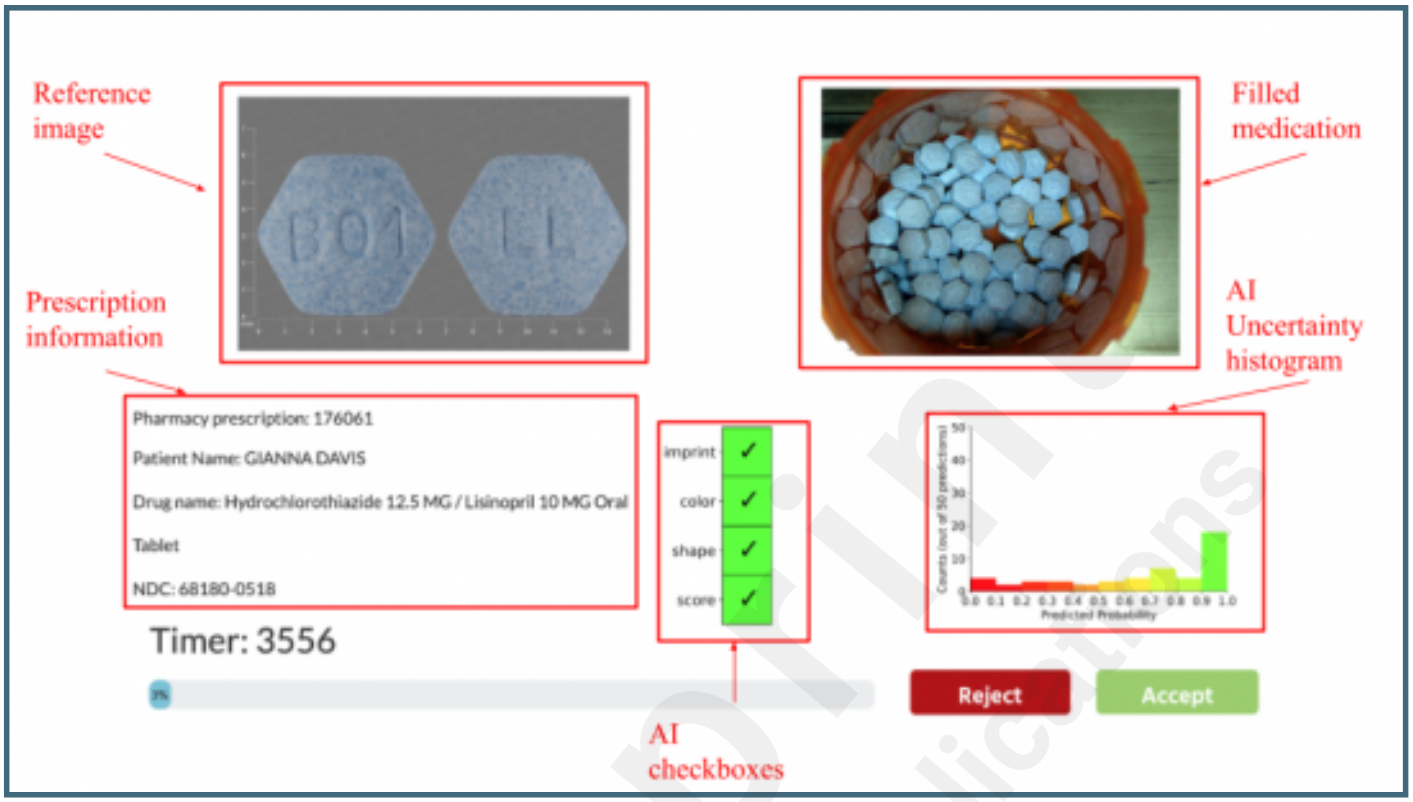
Participant Recruitment Timeline.



Interface for the AI aid without the uncertainty plot. Checkboxes indicate AI's recommendation. When all four checkboxes are green, the AI advises to accept otherwise, to reject.



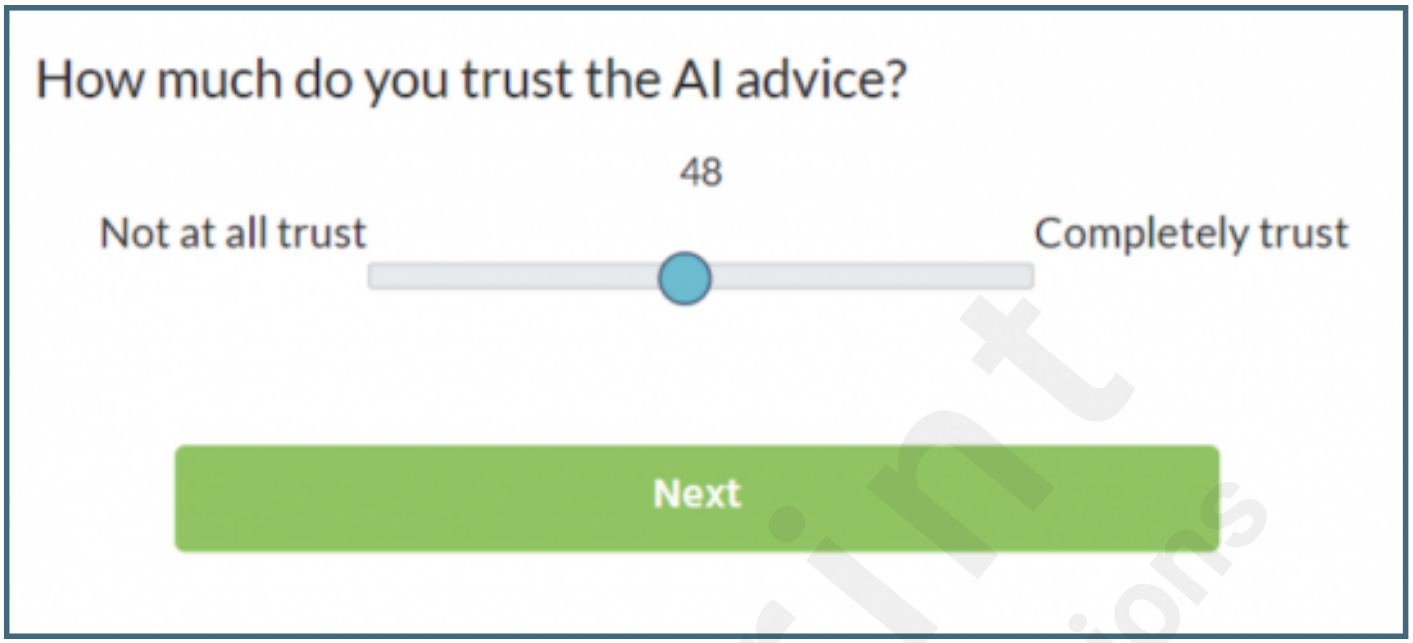
Interface for the AI aid with the uncertainty plot. In addition to Figure 2, the histogram shows the distribution of the 50 predicted probabilities associated with the predicted NDC.



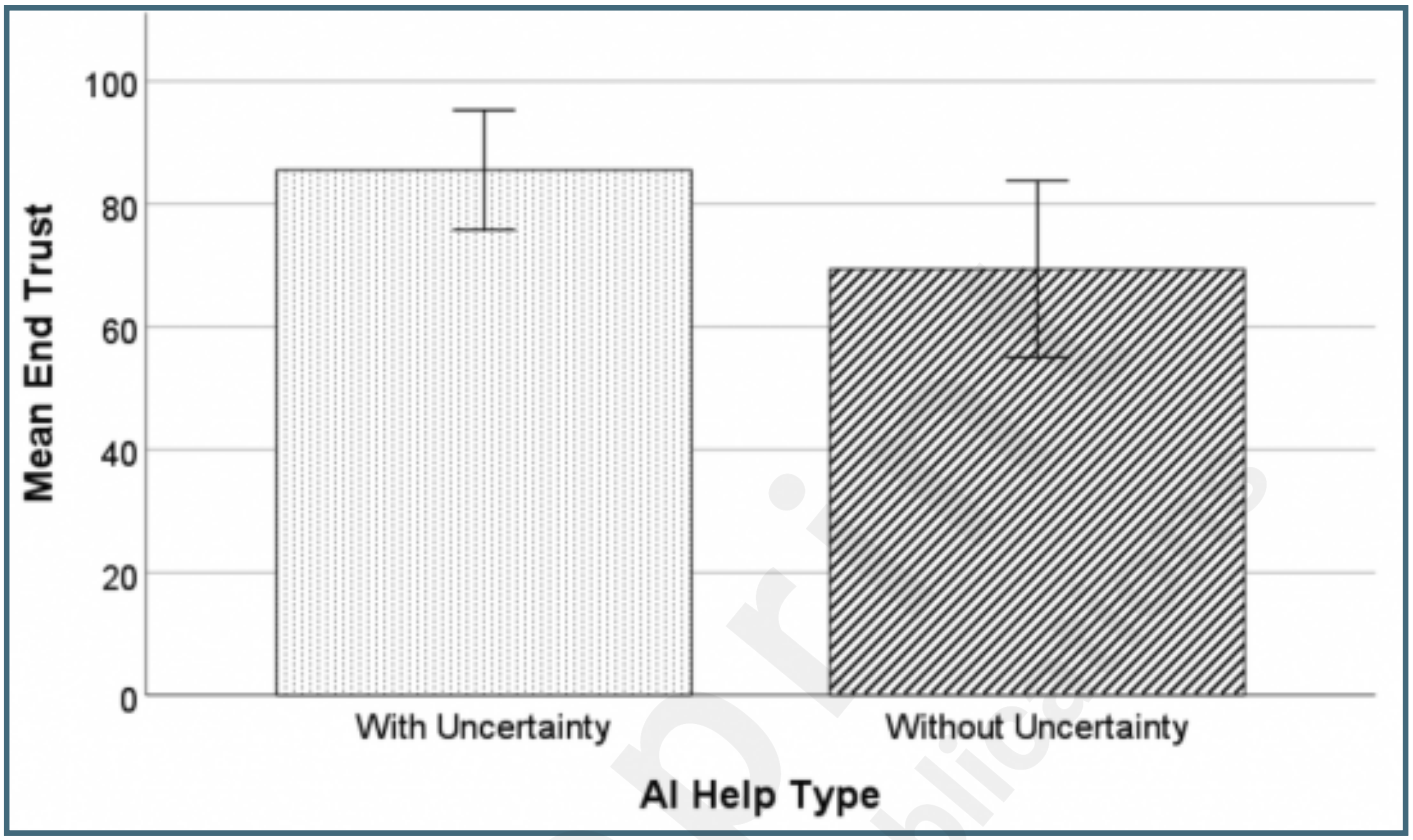
Four potential AI performance patterns according to Signal Detection Theory.

AI Response	State of the World (Fill Image Accuracy): Signal Present (Incorrect Fill)	State of the World (Fill Image Accuracy): No Signal (Correct Fill)
Alert	**Hit** (AI *rejects* the *incorrect* drug)	**False alarm** (AI *rejects* the correct drug)
No Alert	**Miss** (AI approves the *incorrect* drug)	**Correct rejection** (AI approves the correct drug)

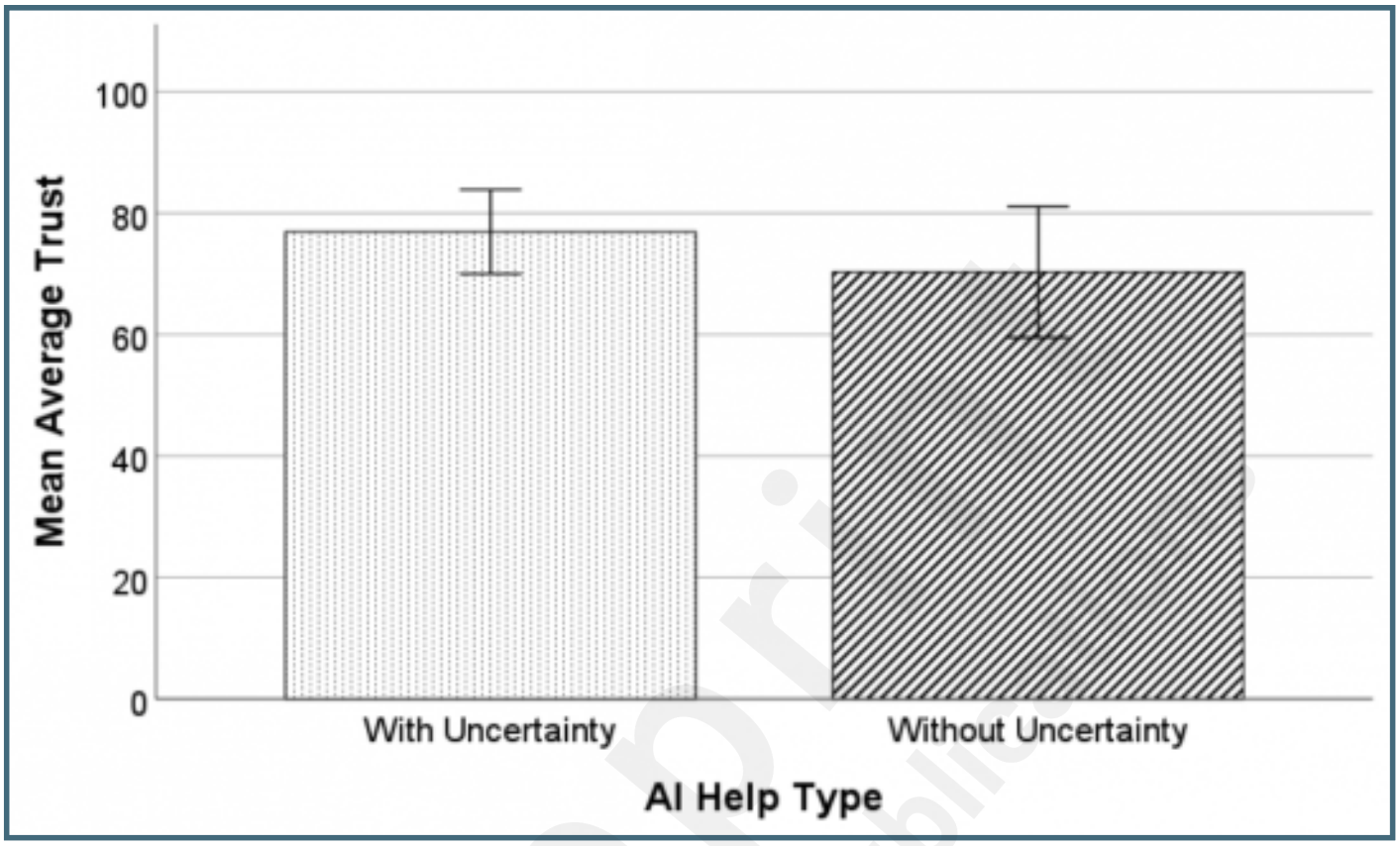
Participants reported their trust in the recognition AI on a visual analog scale.



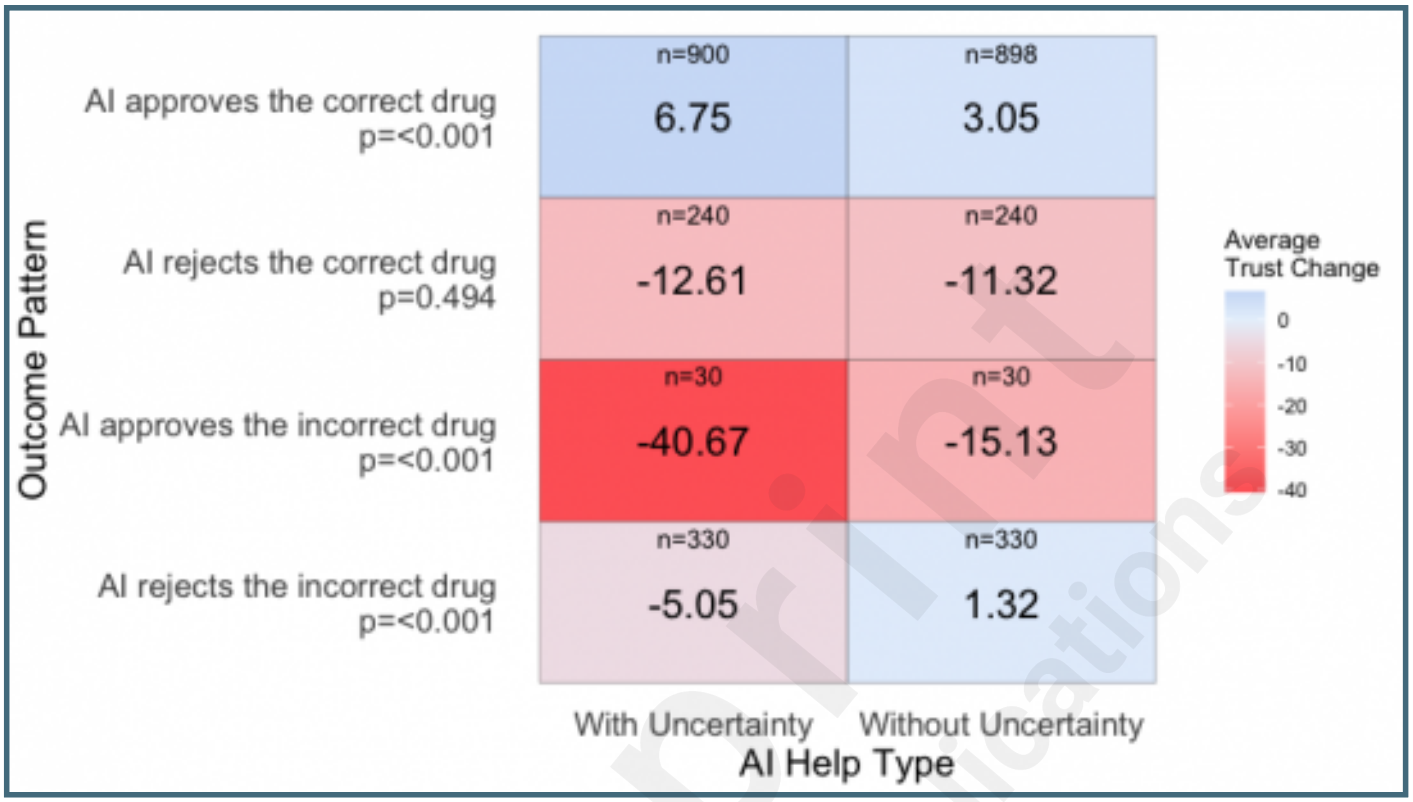
Mean end trust by AI aid with or without the uncertainty plot. The error bars represent +/- 2 standard errors.



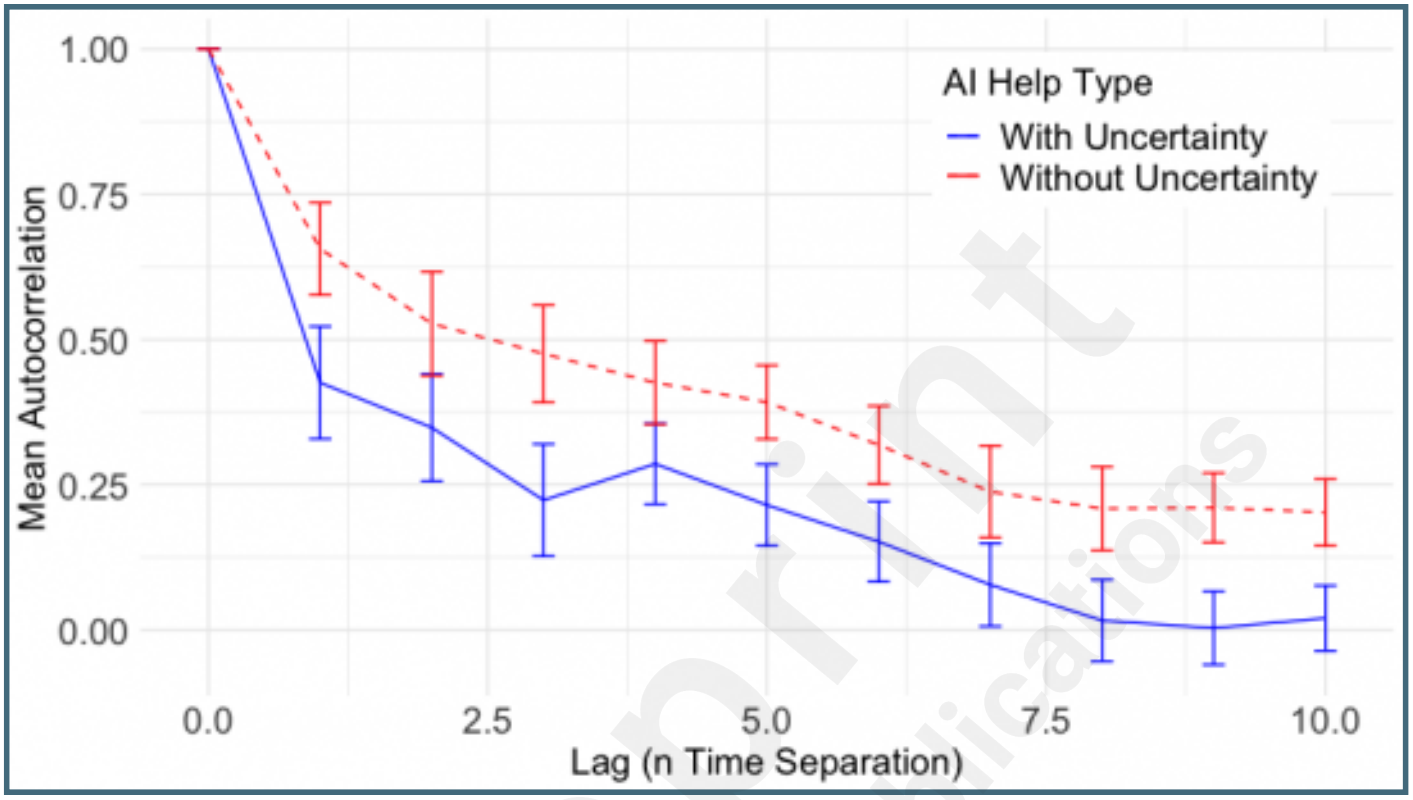
Mean average trust by AI aid with or without the uncertainty plot. The error bars represent +/- 2 standard errors.



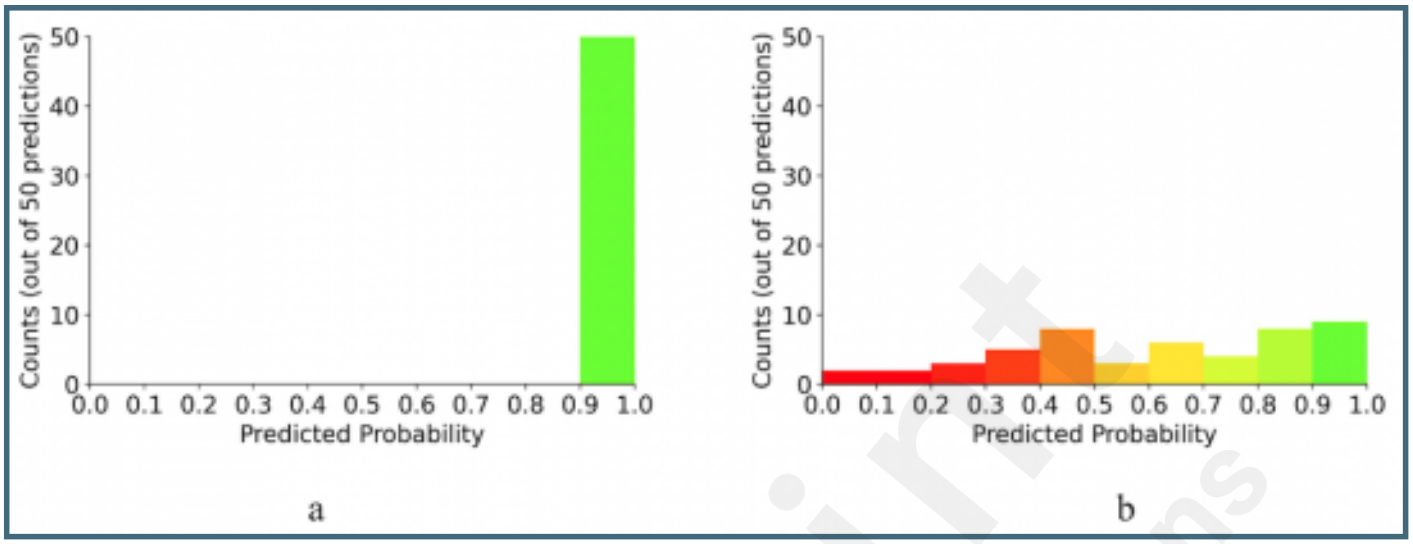
Trust change for different AI performance patterns. Red shades represent negative trust change, and blue shades represent positive trust change.



Autocorrelation of trust as a function of time separation. The blue solid line represents AI aid with the uncertainty plot, and the red dashed line represents the AI aid without the uncertainty plot. The error bars represent +/- 2 standard errors.



Uncertainty plot with narrower (left) and wider (right) interquartile range. Interquartile range is a measure of statistical dispersion, or how spread out the data points are.



Mean interquartile range by outcome pattern. The error bars represent +/- 2 standard errors.