Digestive and Liver Disease xxx (xxxx) xxx

Contents lists available at ScienceDirect

Digestive and Liver Disease

journal homepage: www.elsevier.com/locate/dld

Digestive Endoscopy

Enhancing human-AI collaboration: The case of colonoscopy

Luca Introzzi[a], Joshua Zonca[a,b], Federico Cabitza[c,d], Paolo Cherubini[e], Carlo Reverberi[a,b,*]

[a] *Department of Psychology, Università Milano - Bicocca, Milano, Italy*
[b] *Milan Center for Neuroscience, Università Milano - Bicocca, Milano, Italy*
[c] *Department of Informatics, Systems and Communication, Università Milano - Bicocca, Milano, Italy*
[d] *IRCCS Istituto Ortopedico Galeazzi, Milano, Italy*
[e] *Department of Brain and Behavioral Sciences, Università Statale di Pavia, Pavia, Italy*

ARTICLE INFO

Article history:
Received 3 August 2023
Accepted 23 October 2023
Available online xxx

Keywords:
Artificial intelligence
Cognitive bottlenecks
Cognitive bias
Diagnostic errors
Endoscopy
Hybrid intelligence
Human - AI collaboration

ABSTRACT

Diagnostic errors impact patient health and healthcare costs. Artificial Intelligence (AI) shows promise in mitigating this burden by supporting Medical Doctors in decision-making. However, the mere display of excellent or even superhuman performance by AI in specific tasks does not guarantee a positive impact on medical practice. Effective AI assistance should target the primary causes of human errors and foster effective collaborative decision-making with human experts who remain the ultimate decision-makers. In this narrative review, we apply these principles to the specific scenario of AI assistance during colonoscopy. By unraveling the neurocognitive foundations of the colonoscopy procedure, we identify multiple bottlenecks in perception, attention, and decision-making that contribute to diagnostic errors, shedding light on potential interventions to mitigate them. Furthermore, we explored how existing AI devices fare in clinical practice and whether they achieved an optimal integration with the human decision-maker. We argue that to foster optimal Human-AI collaboration, future research should expand our knowledge of factors influencing AI's impact, establish evidence-based cognitive models, and develop training programs based on them. These efforts will enhance human-AI collaboration, ultimately improving diagnostic accuracy and patient outcomes. The principles illuminated in this review hold more general value, extending their relevance to a wide array of medical procedures and beyond.

As with any other experts, Medical Doctors (MDs) do err. Estimates evaluate diagnostic errors as high as 10–15%, impacting patient health and system costs [1–3]. A promising avenue to reduce such a burden on patients and society is supporting the decision process of MDs with Artificial Intelligence (AI). Here, we evaluate how and why AI may be helpful for the informative setting of colonoscopy, grounding our general conclusions on the fundamental features of human perception and decision-making.

Colonoscopy is at the core of prevention and treatment strategies for colorectal cancer. Ideally, any risky lesion should be removed or reported after a colonoscopy. We know this is not the case: studies using tandem colonoscopies showed that about 25% of adenomas were missed, and 12% of index colonoscopies (the first performed in the tandem procedure) were false negatives. Higher adenoma missing rates are associated with an increased probability of interval cancer [4,5].

Addressing this situation requires systematically examining the *preventable* causes of diagnostic errors, particularly those related to missed detections [6]. Some contributing factors may extend beyond the direct responsibility of MDs, such as organizational constraints (e.g., insufficient procedure time) or patient-related issues (e.g., inadequate preparation). Nonetheless, certain causes are directly linked to the behavior of the MD during the procedure [7,8]. Why may an endoscopist fail to detect a target lesion even in the best conditions? Cognitive Science comes to the rescue to answer this question.

1. Cognitive determinants of errors during colonoscopy

The human brain's processing of sensory information and human decision-making did not evolve to detect lesions flawlessly during standard endoscopy conditions [9]. Human visual search is highly effective when seeking a target stimulus that differs from an irrelevant background based on a single, perceptually salient feature. For example, a red apple in a basket of green apples would be immediately noticed and visually *pop out*. However, the performance drastically changes when the difference between the tar-

* Corresponding author at: Department of Psychology and Milan Center for Neuroscience, Università Milano - Bicocca, Piazza Ateneo Nuovo, 1, 20126 Milano, Italy.
E-mail address: carlo.reverberi@unimib.it (C. Reverberi).

https://doi.org/10.1016/j.dld.2023.10.018
1590-8658/

Fig. 1. Peripheral perception and visual crowding. Focusing on the central fixation cross reveals an intriguing phenomenon: the square on the right side becomes remarkably challenging to perceive. This phenomenon stems from the diminished resolution within the peripheral visual field and the disruptive impact of crowding effects. In stark contrast, the presence of a distinct reddish object on the left immediately captures attention, even under otherwise identical conditions. As an endoscopist's attentional focus is drawn towards the reddish entity, effortless, successful identification of its relevance ensues.

get and background extends over multiple perceptual dimensions that only collectively allow to discriminate the objects. Red apples do not pop out in a basket with red peppers and green apples: you must pay attention to both color and shape because no single property separates the targets from the irrelevant stimuli. In such scenarios, we engage in an intentional exploration of each image patch, patiently searching for one or a few target stimuli. Our search becomes *guided* by a target template: Targets are defined by a multidimensional combination of features that segregate them from irrelevant stimuli or the background [10]. Identifying lesions amidst healthy mucosa is an excellent example of a guided search. Guided search enables us to search for complex patterns at the expense of drastic decreases in speed and accuracy. The guided search can even overlook relevant but untargeted objects, producing astonishing phenomena like inattentional blindness [11,12]. Guided search is also more vulnerable to vigilance decrements and attentional lapses, which tend to occur with increasing time on task, the number of simultaneous tasks, and the operator's mental and physical fatigue [9,13–18].

In addition, without us noticing, detection ability degrades on the periphery of the visual field (Fig. 1) because of reduced visual acuity and crowding [19,20]. Complex objects can only be identified within the *useful field of view*, which is roughly 2.5° centered on the fovea [9,19], equivalent to a 4.4 cm width when looking at a screen 1 m away. To compensate for this limitation, individuals may systematically scan through the image, moving their useful field of view around to cover all locations. However, human memory for already sampled locations proved to be poor, resulting in both less efficient and defective search behavior: less efficient because they may look multiple times at the same location, defective because they may miss a significant proportion of the image [19,21,22].

Higher-order cognitive tendencies, such as the ones involved in misjudgments of probabilities (miscalibration [23,24]), can also impact detection. For instance, the *over-alternation bias* leads us to believe that if something has just occurred, it is less likely to occur again [25,26]. This tendency may contribute to the observed increase in misses following detecting the first polyp during a colonoscopy [7,27–29].

Taken together, these perceptual and cognitive bottlenecks indicate that human experts may face challenges in detecting certain endoscopic targets, even in high-resolution still images (Fig. 2) [11,30]. Live endoscopy adds further difficulties to the examination of still images. Endoscopists not only need to visually search for lesions but also navigate while keeping track of the probe's position and ensuring adequate mucosa exposure. These tasks are challenging on their own and become even more arduous when performed simultaneously. *Dual-task* conditions are known to decrease performance in both laboratory settings [31] and medical settings [32–35]. Although training can reduce interference between the tasks, the implementation of non-standard procedures, such as demanding maneuvers, can lead to detrimental dual-task interactions and compromise lesion detection abilities [36].

Finally, once a suspicious lesion is detected, the endoscopist must make an optical diagnosis to determine its nature. While the human brain excels at pattern recognition (especially after training), this task can sometimes overwhelm our capacities. Unfortunately, humans often underestimate the susceptibility to errors in certain tasks and tend to be overconfident, reporting higher accuracy estimates than they actually achieve. For example, in a classic study before right heart catheterization, physicians estimated values for pulmonary capillary wedge pressure, cardiac index, and systemic vascular resistance while stating their confidence in their estimates. Accuracy was generally low and did not increase with experience. However, confidence increased with experience [37]. Overconfidence has been documented among MDs (and other experts), and it is a common occurrence. Overconfidence may cause fixation [38] and hinder learning and correction, particularly when physicians do not receive systematic feedback on highly confident yet incorrect judgments [39]. For example, an endoscopist highly confident that a removed polyp is an adenoma may decide not to access the histology that shows it was not.

2. The role of AI tools in colonoscopy: promises and dangers

Following the analysis of the cognitive intricacies of endoscopists' performance, it becomes evident that they could benefit from assistance. An ideal candidate for support should address, if not all, at least some of the aforementioned cognitive bottlenecks.

Modern AI tools based on, e.g., deep convolutional neural networks, emerge as a promising option on paper. These tools can search for multiple complex visual targets without succumbing to attentional bottlenecks and without experiencing performance degradation in the periphery of the visual field. Moreover, AI exhibits strong pattern recognition capabilities, relying on different features and mechanisms than humans, thereby likely failing for reasons distinct from those affecting human performance [40]. Such partial divergence further improves the utility of AI, reducing the chances that both humans and AI fail on the same item. Overall, AI performance in lesion detection has been reported as similar to the best human experts [41,42]. Leveraging its capacities, AI can be harnessed to assist MDs in detecting lesions, thus transforming potentially missable lesions (e.g., a small flat lesion briefly glimpsed in the periphery of the visual field) into unmistakable visual targets. For example, AI could surround the lesion with a color absent from the colon color palette, making it pop out (Fig. 2).

The potential benefits of AI assistance in endoscopy extend beyond the detection of exposed lesions. By partially relieving endoscopists from the burden of visual search, AI could liberate cognitive resources for other human-exclusive tasks, such as navigation, framing optimization, and tracking mucosa exploration. This, in turn, can further enhance the overall procedure's efficiency and accuracy. After detection, AI can assist the endoscopist in the optical diagnosis of lesions by providing an additional online opinion on the nature of the lesion. Importantly, the AI opinion is not infallible but would provide a further source of evidence for the final decision of the MD, as in double-reading or second-opinion scenarios.

Finally, AI assistance may provide a valuable feedback tool for junior endoscopists. The availability of feedback across multiple examinations plays a pivotal role in fostering a successful learning process. By receiving AI-supported feedback, junior endoscopists can gain valuable insights, learn from their mistakes, and refine

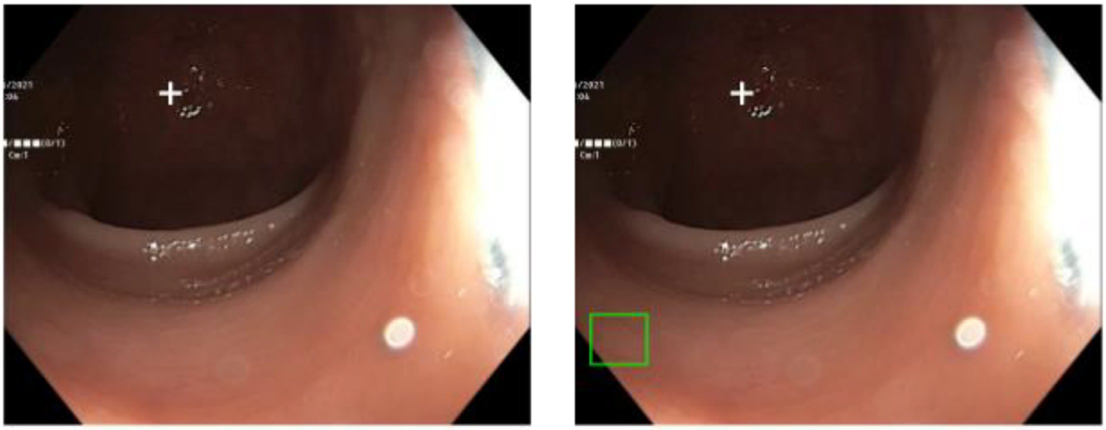

Fig. 2. Pop-out effect during colonoscopy. Left panel: Colonoscopy requires a visual search of lesions while navigating the colon. In this example, we assume that the endoscopist is fixating at the center of the lumen (fixation cross) while scanning the mucosa. In typical endoscopic conditions (80 cm large display, 1 m away from the endoscopist), a lesion similar to the surrounding mucosa and 17° away from the fixation point might be easily missed. Right panel: AI can assist detection by making an object with low visual salience into a highly distinctive green square. This effectively changes the visual search required: from a slow error-prone serial search to a fast, effortless, highly accurate parallel search. Once the endoscopist's attention has been attracted to the suspect lesion, she can proceed with an optical diagnosis.

their skills, ultimately leading to enhanced performance and diagnostic accuracy over time [43].

So, all set? Is the use of AI in colonoscopies already an obvious choice today that would certainly improve the procedure? Should the technical improvement of AI performance be considered the only remaining concern? Not so fast. While AI holds great promise, it is important to recognize that the final performance of a human expert assisted by AI depends not only on the competence of the AI and the human but also, critically, on the quality of the interaction between the two.

To better understand this, we can conceptualize AI assistance as that of a human expert who evaluates similar evidence and provides advice but leaves the final decision to the collaborator. Optimal integration of opinions, whether from human or artificial agents, should consider the following factors: their judgments, their confidence in the correctness of their opinions, and the average objective accuracy of the agents in similar situations in the past. Having this information, one could appropriately weigh the two agents' opinions, crediting more influence to the most reliable one [44–48]. If both opinions appear equally reliable, a rational decision would be to integrate them with equal weight, based on the statistical principle that aggregating imperfect judgments reduces error [49]. For instance, consider a scenario where you and your friend attended a ceremony and are now estimating the number of people present. You estimate "about 50" and your friend says "about 70". If you trust her judgment as much as yours, the most likely count would be an integrated estimate of "about 60". However, if she adds, "I am pretty sure! I counted them, by and large", while you did not, you should trust her judgment on this specific occasion more than yours, leading to a final best guess closer to 70 than 50. Even in these ideal conditions, where all relevant information is available, humans deviate from an optimal integration of opinions [50,51].

Deviations from optimal integration become even more common when some information is missing. For instance, one may not know how accurate collaborators are on average, or collaborators may communicate their opinions but not their confidence level in the advice. In such cases, decision-makers must rely on estimates, assessing whether the collaborator is more or less accurate in general and how confident she appears in the specific case. Unfortunately, humans tend to err in these estimates, often overestimating their chance of being correct, leading to an implicit, unwarranted preference for their own opinions, an *ego bias* [52–57]. In the earlier example, each friend may systematically trust their judgment a little more than the other, resulting in two distinct opinions: "No matter what she says, I know better: they were close to 50″, and "No matter what he says, they were close to 70″. The ego bias can inhibit a calibrated use of others' advice, generating negative impacts on collective decisions: extensive evidence has shown that aggregating informed opinions improves effectiveness [50,58], promoting the *wisdom of crowds* [59,60]. The ego bias can be reduced by providing additional information on the competence and expertise of the collaborator (i.e., the advisor), which can signal an increased probability of high-quality advice [52].

When the partner is an AI, we witness the same biases observed in the interaction with a human collaborator (e.g., ego bias), possibly exacerbated by the specificities of the human-AI interaction (Table 2). The communication layer between AI and humans may be suboptimal due to design flaws. More interestingly, communication failure can also stem from a deeper inability to understand how AI makes decisions and how likely AI may err in specific conditions [61]. Humans spontaneously use introspection and intuitive knowledge of how the human mind feels and reason, building a naive "theory of (human) mind" [62]. This theory of mind makes it easier to infer how other human experts might have reached a different conclusion and weigh the likelihood that they are correct. By contrast, modern AI learns and reaches conclusions in a different, non-human way. Critically, the opaqueness of the AI's internal computational processes may decrease calibrated reliance on the system [63,64]. The lack of transparency (i.e., intelligibility and explainability) of AI may lead to different suboptimal outcomes. First, it may prevent the formation of reliable case-by-case expectations of AI accuracy [65,66]. For example, a human operator may be unable to understand why an AI assistant made a mistake, hindering the ability to predict when errors are more likely to occur in the future. Second, it can induce high variability in the a-priori tendencies toward the evaluation of AI competence, leading to general trust (algorithm appreciation) or general mistrust (algorithm aversion) towards AI technology [67–70]. Both these tendencies may result in suboptimal exploitation of information provided by AI, giving too much or too little weight to AI's outputs. Third, the lack of transparency may lead humans to overreact to AI errors: if humans cannot understand the cause of an AI error or if they consider it anomalous compared to typical human errors, they may be tempted to infer that the AI is generally unreliable and incompetent, leading to mistrust and misuse of the AI system itself [71,72].

Table 1
Cognitive bottlenecks that can hinder the endoscopist's performance in a colonoscopy and how AI could help overcome them. The first four rows refer to features of the attentional processes, mainly relevant for detection. The last two rows refer to judgment and decision-making processes, which are more relevant for optical diagnosis.

Cognitive process *(domain)*	Bottleneck	AI support
Visual search *(orienting of visual attention)*	Targets defined by a combination of more than one perceptual feature shared by other objects in the scene require lengthy serial search.	Framing potential polyps with a distinctive color turns a serial search into an easy "pop-out" search.
Inhibition of return *(orienting of visual attention)*	When a location in space is explored (even for a very short period of time: 0.3 s is enough), and no targets are found, that location gets an automatic "inhibitory tag": gaze is impeded from spontaneously returning there.	AI is not affected by IOR. If a human fails to detect a polyp in a location, the appearance of a distinctive-color frame in that location automatically returns attention there, overcoming IOR and allowing for a "second guess" before the lesion disappears from view.
Cognitive overload, dual task, multi-tasking *(mental workload)*	When two or more non-automatized tasks are performed simultaneously, they interfere even if they do not require the same effectors (e.g., one task requires moving hands, the other requires looking): performance in one or both tasks deteriorates.	Colonoscopy is a multitasking procedure: navigation, detection, and categorization. Each task, if not fully automatic, drains cognitive resources, increasing the chances of errors in other tasks. AI, by supporting detection and categorization, frees resources for navigation.
Vigilance decrement, attention lapses, fatigue *(sustained attention)*	For all tasks involving attention, the number of errors increases with time-on-task. This is unavoidable and includes both perceptual errors (misses) and cognitive errors (wrong decisions). The attention lapses are even more probable when the rate of true targets is low: e.g., the fewer polyps to be detected there are in a long colonoscopy, the more probable is that one of them is missed	AI support can shorten the overall time-on-task, decreasing the risk of attention lapses. Furthermore, target-framing captures human attention even if the operator is momentarily in a lapse.
Miscalibration of diagnostic probability *(judgment and decision-making)*	Even expert judgments may overestimate *or* underestimate diagnostic probabilities. These effects can decrease human accuracy in optical diagnosis. This miscalibration is decreased by increased information, and spending more time thinking about a judgment	AI diagnosis, even if not faultless, is a "second opinion" that forces the human to compare and weigh the two diagnoses, possibly reducing the adverse effects of miscalibration.
Overconfidence or illusion of validity *(judgment and decision-making)*	People overestimate the probability that their judgments are correct.	AI diagnosis, when not convergent with the human diagnosis, can induce second thoughts and possibly reduce the effects of overconfidence

Table 2
How much to trust a message from an agent that has not perfect reliability (i.e., accuracy) depends on the agent's reliability (which can be further specified into sensitivity and specificity) and on how much the agent is confident (how much the agent is calibrated in its informed guesses about the correctness of its messages). Information about AI confidence and reliability is needed, but humans should also be trained on how to use it appropriately.

How much should I trust the AI opinion?	
Problem	Solution
Human lacks info on AI Confidence & Reliability (C&R)	Provide high-quality information on AI C&R
AI C&R is available but not used	Improve AI design: Develop human-friendly formats for embedding info on AI C&R into the current stimulus response output
AI C&R available in the appropriate format, but human is overconfident (e.g., ego bias, algorithm aversion, fixation)	Structured training. Better human-AI collaboration protocols. Humans must be able to build a mental model of how AI works.
AI C&R is available in the appropriate format, but human is underconfident (e.g., automation bias)	

One of the possible solutions to mitigate under-reliance on AI due to lack of transparency is allowing an AI to communicate its confidence in the current response output, which has been shown to help the calibration of trust in the system and reliance on its advice [73–75].

3. Human-AI interaction in colonoscopy: empirical evidence

The previous section elucidated general principles and challenges related to achieving an effective human-AI interaction. Colonoscopy presents an intriguing setting for applying these principles.

3.1. Human-AI interaction in lesion detection (CADe)

Available AI devices exhibit very high sensitivity and overall performance on par with or better than top human experts [42,76]. Support from an AI tool with high sensitivity should alleviate the cognitive bottlenecks related to orienting attention, sustained attention, and dual-task workload (Table 1). How could dysfunctionalities and suboptimal outcomes show up in this context? A primary source would likely stem from a suboptimal strategic decision, i.e., MDs not delegating to the AI a substantial part of the detection task (i.e., under-use), thus not exploiting the potentialities of relieving dual-task cognitive requirements and freeing resources for other human-only tasks such as navigation. Besides, processing false positives detected by AI may drain too much attention and time if each triggers a thorough evaluation. It would be important to identify simple cues for swiftly segregating AI outputs into likely true and likely false to counter this potential issue. The temporal persistence of the AI signal [77] appears to be a promising candidate, as a longer persistence (e.g., 200 ms) is associated with a higher likelihood of the AI signal corresponding to a true lesion.

The characteristics of available AI tools for colonoscopy, the challenges faced by endoscopists, and our understanding of human cognition collectively support strong expectations for a beneficial role of AI. However, we also raised some cautionary notes on how human-AI interaction could go wrong. Which side does the empirical evidence favor? The randomized controlled trials conducted to date evaluate the effectiveness of CADe devices. Overall, the

Table 3
Reviews with metanalysis on Randomized Controlled Trials in which AI-assisted colonoscopy has been compared to standard non-AI assisted colonoscopy.

Authors	Title	Number of studies finally included	Total number of participants
Deliwala et al. (2021) [101]	Artificial intelligence (AI) real-time detection vs. routine colonoscopy for colorectal neoplasia: a meta-analysis and trial sequential analysis	6	4996
Hassan et al. (2021) [102]	Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis	5	4354
Huang et al. (2022) [83]	Effect of artificial intelligence-aided colonoscopy for adenoma and polyp detection: a meta-analysis of randomized clinical trials	10	6629
Mori et al. (2023) [103]	Impact of Artificial Intelligence on Colonoscopy Surveillance After Polyp Removal: A Pooled Analysis of Randomized Trials	9	5796
Spadaccini et al. (2021) [104]	Computer-aided detection versus advanced imaging for detection of colorectal neoplasia: a systematic review and network meta-analysis	50	34,445
Zhang et al. (2021) [105]	Artificial Intelligence-Aided Colonoscopy for Polyp Detection: A Systematic Review and Meta-Analysis of Randomized Clinical Trials	7	5427

evidence strongly supports the adoption of AI for improving lesion detection during colonoscopy (Tables 3 and 4) [76,78–81]. Notably, a recent review revealed that CADe systems reduced the adenoma miss rate by an impressive 65% in colonoscopies [76]. This means that any potential factors in human-AI interaction that could be detrimental to CADe's effectiveness, if present, have been largely surpassed by the advantages it offers.

Evidence also suggests a potential differential impact of AI on lesions based on their morphology and size, aligning with psychological expectations. Small and flat lesions, less visually salient, are expected to result in a higher miss rate. Tandem studies have confirmed these expectations, as small and flat morphology has negatively impacted on the detection rate [4]. CADe systems prove particularly beneficial in aiding the detection of these challenging lesions [76,82,83]. Nevertheless, larger (> 10 mm) or polypoid adenomas, which are more apparent, also benefit from CADe assistance, albeit possibly to a lesser extent [76,84]. If these patterns are further corroborated in future studies, they could shed light on the underlying mechanisms behind CADe's effectiveness. A specific advantage of CADe for perceptually "difficult" lesions could be associated with relieving perceptual bottlenecks (Table 1). In contrast, a more general benefit across lesion types might be attributed to improved sustained attention and relief of dual-task burden, enabling better navigation. The role of CADe assistance in mitigating sustained attention lapses is further underscored in a study considering endoscopists' performance during the day [85]. It was found that the number of endoscopies performed in a day by an endoscopist negatively impacted performance [86–88], but the use of CADe eliminates this adverse effect [85].

Finally, and importantly, a study evaluated the potentially damaging effect of AI false positives on endoscopists' performance. While false positives are not rare occurrences (e.g., ∼27 per colonoscopy), human experts can readily recognize them as such without any delay in the colonoscopic procedure in most cases. In other words, human experts in the study applied effective rules of thumb for ignoring them without in-depth examination [77]. However, mastering this remarkable ability may require appropriate training.

While reviews and meta-analyses provide robust support for the widespread adoption of AI in colonoscopy, a few studies have reported no improvement in quality metrics for in vivo endoscopy with versus without AI assistance [89–91].

Two of the studies were conducted in "real-world settings", which involved retrospective data analysis from medical centers that introduced AI support tools [90,91]. Although this approach lacks the rigor of randomized controlled trials (RCTs), it could offer valuable insights. These negative results represent a precious opportunity for study, allowing us to understand what, in the human-machine interaction, turned a documented useful AI tool into a failed assistant.

A hypothesis shared by both studies refers to a strategic error in approaching AI. The studies report that MDs may have developed "a sense of false confidence and overreliance on AI technology". In other words, they may have attributed unrealistic capabilities to the AI, leading to a misjudgment of the respective roles of the human and the artificial agent. This could have resulted in "unconscious degradation in the quality of mucosal exposure" or a "somewhat less scrupulous performance". Consistent with this hypothesis, Levy and collaborators report faster colonoscopic procedures in the CADe group [91]. This strategic error might have been facilitated by the limited training received on the optimal use of the new AI tool. For instance, Ladabaum and colleagues decided to make "CADe available without any interventions beyond encouragement and basic start-up training."

These negative results offer important warnings regarding using AI tools in clinical settings. Although AI tools demonstrated outstanding standalone detection abilities, they might not necessarily provide a measurable advantage in real-world clinical scenarios. The studies suggest that strategic misunderstandings, possibly due to inadequate training and poor human-AI fit, could have contributed to this outcome. However, there are still many aspects to be comprehended fully. The raised hypotheses must be formally tested, and other factors should be explored, such as the baseline performance of medical professionals, their attitude towards technology, trust towards AI, understanding of AI operations, the interaction protocols (e.g., either human-first or AI-first), and the specific features of the tools available in the facility. Considering the key role of successful human-AI collaboration in delivering promised clinical outcomes, future studies should include incorporating a more comprehensive set of measurements that capture both the human and technological context in which AI is introduced [66,92,93].

3.2. Human-AI interaction in lesion characterization (CADx)

Unlike CADe, CADx advice is provided to endoscopists at a moment during the colonoscopy when dual-task requirements are reduced. Navigation and detection are paused, allowing endoscopists to concentrate solely on diagnosing the detected lesion, thus avoiding the potential dangers of a dual-task situation or attentional

Table 4
Studies on AI-assisted colonoscopy for detection (CADe) and characterization (CADx) of colon lesions.

Authors	Title	Sample size	Study type	Procedure
Ahmad et al. (2022) [89]	Evaluation of a real-time computer-aided polyp detection system during screening colonoscopy: AI-DETECT study	658	RCT	CADe
Barua et al. (2022) [94]	Real-Time Artificial Intelligence–Based Optical Diagnosis of Neoplastic Polyps during Colonoscopy	518	clinical study	CADx
Gimeno-García et al. (2023) [82]	Usefulness of a novel computer-aided detection system for colorectal neoplasia: a randomized controlled trial	370	RCT	CADe
Glissen Brown et al. (2022) [106]	Deep Learning Computer-aided Polyp Detection Reduces Adenoma Miss Rate: A United States Multi-center Randomized Tandem Colonoscopy Study (CADeT-CS Trial)	232	RCT	CADe
Gong et al. (2020) [107]	Detection of colorectal adenomas with a real-time computer-aided system ENDOANGEL): a randomised controlled study	704	RCT	CADe
Hassan et al. (2023) [95]	Comparative Performance of Artificial Intelligence Optical Diagnosis Systems for Leaving in Situ Colorectal Polyps	176	Prospective comparison trial	CADx
Hüneburg et al. (2023) [108]	Real-time use of artificial intelligence (CADEYE) in colorectal cancer surveillance of patients with Lynch syndrome-A randomized controlled pilot trial (CADLY)	101	RCT	CADe
Kamba et al. (2021) [109]	Reducing adenoma miss rate of colonoscopy assisted by artificial intelligence: a multicenter randomized controlled trial	358	RCT	CADe
Karsenti et al. (2023) [110]	Effect of real-time computer-aided detection of colorectal adenoma in routine colonoscopy (COLO-GENIUS): a single-center randomised controlled trial	2592	RCT	CADe
Ladabaum et al. (2023) [90]	Computer-aided Detection of Polyps Does Not Improve Colonoscopist Performance in a Pragmatic Implementation Trial	1008	retrospective observational study	CADe
Levy et al. (2022) [91]	Artificial Intelligence-Aided Colonoscopy Does Not Increase Adenoma Detection Rate in Routine Clinical Practice	4414	retrospective observational study	CADe
Liu et al. (2020) [111]	Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy	1026	RCT	CADe
Lu et al. (2023) [85]	Assessment of the Role of Artificial Intelligence in the Association Between Time of Day and Colonoscopy Quality	1780	RCT	CADe
Luo et al. (2021) [112]	Artificial Intelligence-Assisted Colonoscopy for Detection of Colon Polyps: a Prospective, Randomized Cohort Study	150	RCT	CADe
Lux et al. (2022) [113]	Pilot study of a new freely available computer-aided polyp detection system in clinical practice	41	Pilot study	CADe
Nakashima et al. (2023) [114]	Clinical Evaluation of Computer-Aided Colorectal Neoplasia Detection Using a Novel Endoscopic Artificial Intelligence: A Single-Center Randomized Controlled Trial	415	RCT	CADe
Quan et al. (2022) [115]	Clinical evaluation of a real-time artificial intelligence-based polyp detection system: a US multi-center pilot study	600	Pilot study	CADe
Repici et al. (2020) [79]	Efficacy of Real-Time Computer-Aided Detection of Colorectal Neoplasia in a Randomized Trial	685	RCT	CADe
Repici et al. (2022)[116]	Artificial intelligence and colonoscopy experience: lessons from two randomised trials	660	RCT	CADe
Reverberi et al. (2022) [66]	Experimental evidence of effective human-AI collaboration in medical decision-making	504	Experimental study	CADx
Rondonotti et al. (2022) [96]	Artificial intelligence-assisted optical diagnosis for the resect-and-discard strategy in clinical practice: The Artificial intelligence BLI Characterization (ABC) study	389	Prospective cohort study	CADx
Shaukat et al. (2022) [117]	Computer-Aided Detection Improves Adenomas per Colonoscopy for Screening and Surveillance Colonoscopy: A Randomized Trial	1440	RCT	CADe
Su et al. (2020) [118]	Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos)	659	RCT	CADe
Wallace et al. (2022) [84]	Impact of Artificial Intelligence on Miss Rate of Colorectal Neoplasia	230	RCT	CADe
P. Wang, Liu, Berzin, et al. (2020) [80]	Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study	1046	RCT	CADe
P. Wang, Liu, Glissen Brown, et al. (2020) [119]	Lower Adenoma Miss Rate of Computer-Aided Detection-Assisted Colonoscopy vs Routine White-Light Colonoscopy in a Prospective Tandem Study	185	RCT	CADe
Wei et al. (2023) [120]	Evaluation of Computer-Aided Detection During Colonoscopy in the Community (AI-SEE): A Multicenter Randomized Clinical Trial	769	RCT	CADe
H. Xu et al. (2023) [81]	Artificial Intelligence-Assisted Colonoscopy for Colorectal Cancer Screening: A Multicenter Randomized Controlled Trial	2352	RCT	CADe
L. Xu et al. (2021) [121]	Artificial intelligence-assisted colonoscopy: A prospective, multicenter, randomized controlled trial of polyp detection	3059	RCT	CADe
Yamaguchi et al. (2023) [122]	Impact of an artificial intelligence-aided endoscopic diagnosis system on improving endoscopy quality for trainees in colonoscopy: Prospective, randomized, multicenter study	231	RCT	CADe
Yao et al. (2022) [123]	Effect of an artificial intelligence-based quality improvement system on efficacy of a computer-aided detection system in colonoscopy: a four-group parallel study	1076	RCT	CADe

slips. Furthermore, the performance of available CADx systems is comparable, if not inferior, to that of experienced endoscopists [66,94–96]. These facts raise the possibility that assistance would even deteriorate the performance of the best endoscopists. In this context, potential sources of suboptimal human-AI collaboration could include:

- Overreliance: The MD relies too much on the AI's opinion and does not critically evaluate its fallible output, thus exhibiting *automation bias*.
- Underreliance: The MD underestimates the AI's capabilities and does not take into consideration its informative suggestions, thus exhibiting *algorithmic aversion*.
- Lack of transparency: When differing opinions emerge, the MD cannot determine who is most likely correct, indicating a failure to estimate the relative (human/AI) confidence in the judgment.

Reverberi and collaborators [66] investigated the influence of AI opinion on the optical diagnosis of both expert and non-expert endoscopists using an offline repeated-judgment design. The study revealed that the AI output substantially influenced both groups of endoscopists. Crucially, all endoscopists were more likely to accept AI advice when it was correct than when it was wrong. This remarkable ability to discern the quality of AI opinion significantly improved the integrated human-AI decision accuracy, especially among non-experts. The effect on experts was smaller, but it still held significance thanks to the study's high power. How could the endoscopist tell that the advice of AI was better or worse than their own? Endoscopists generated an intuitive assessment of AI opinion's quality on a lesion-by-lesion basis, likely based on subtle features of the AI output like its temporal stability. Such intuitive assessment predicted AI accuracy on the specific lesion, thus allowing endoscopists to accept only good advice [66].

Initial studies evaluating the clinical utility of CADx during colonoscopy procedures largely support these conclusions (see Table 4 for a list of significant CADx studies). Results show that experienced endoscopists' optical diagnosis performance did not significantly benefit from AI assistance, and no harmful effects occurred even when AI performance was inferior to human experts [95,96]. However, one study could not demonstrate a significant increase in sensitivity and specificity, even among non-experienced endoscopists [94], underscoring the need for further research to understand the factors influencing CADx's variable impact.

4. Enhancing human-AI collaboration

Recognizing the critical importance of an effective human-AI collaboration, we must address the question: How can we enhance the interaction between medical professionals (MDs) and AI devices? Several key areas warrant attention to facilitate optimal human-AI integration in medical settings.

First and foremost, establishing an evidence-based cognitive model for human-AI interaction is imperative, as this would allow for designing the most effective human-AI collaboration protocols [68,97,98]. While existing insights provide valuable perspectives, they remain preliminary, and a comprehensive understanding of the cognitive bottlenecks (Table 1) leading to performance disparities is lacking. Further in-depth research is needed to identify the most significant impediments, determine the effectiveness of various improvement mechanisms, and ascertain the impact across different contexts (e.g., hospital vs. community) and individuals.

Second, developing short training programs is essential to help users maximize the potential of new AI devices through appropriate general strategies. Stimulating in users the build-up of an intuitive "theory of the artificial mind" involving a rich internal model of its distinct workings compared to human cognition could enable better predictions of AI accuracy and reliability [65]. Such understanding will help identify situations where AI may be prone to errors or exceptionally dependable and improve the human-AI fit. More generally, the study of how machines interpret medical findings and present them to the human interpreter, called *machine semiotics*, has been proposed as a core competence for future MDs beyond gastroenterology [99].

Third, continuous monitoring of outcomes in different settings is crucial, along with a quantitative assessment of performance beyond mere accuracy [93] and the meticulous reporting of any AI failures for subsequent adjustments in output design or for devising targeted training.

Fourth, human variability should be considered. Our perceptual, attentional, and navigation skills or ability to detect patterns are unequal, as is our level of expertise in a procedure. Therefore, it is important to recognize that AI devices may benefit MDs differently and should adapt to their attitudes, skills, and profiles. For some, AI could offer minor improvements, while for others, it might be as invaluable as a pair of glasses: For example, we observed that expert endoscopists benefit less from AI support for optical diagnosis. Other individual parameters may be at play, such as susceptibility to dual-task performance. Additionally, individual attitudes, like openness to technology and readiness to adoption, can influence the acceptance of AI. Some people may be biased against using relatively new technical solutions, even if proven advantageous for a specific task, due to fears of long-term unknown effects [100]. Recent discussions about the risks of AI may drive more people to close themselves off to AI support systems prejudicially.

5. Conclusions

The use of colonoscopy for screening, diagnosis, and intervention has proven highly effective for colorectal cancer prevention. However, the procedure remains susceptible to errors, prompting significant efforts to reduce them [7,8]. The introduction of AI-based tools to assist MDs during this complex procedure has garnered considerable interest, even though some negative observations also started to emerge [92].

In this context, cognitive science may provide an important contribution, unraveling the neurocognitive foundations of diagnostic errors observed in clinical practice and thus contributing to a theory-informed and evidence-based design of decision support systems. By identifying the multiple bottlenecks in perception, attention, and decision-making that contribute to diagnostic errors, cognitive science provides crucial insights for designing effective interventions, thus leading to the optimal integration of AI systems into clinical practice. Moreover, it provides a conceptual framework for comprehending the potentialities and pitfalls of human-AI collaboration, paving the way for its enhancement.

Colonoscopy serves as an insightful application domain and clinical setting due to a specific task situation in which a fruitful collaboration between MDs and AI can be easily envisaged and due to the acute awareness of the medical community on the importance of addressing diagnostic errors. However, the principles illuminated in this case hold a more general value, extending their relevance to a wide array of medical procedures and beyond.

Conflict of interest

C.R. is offering paid advice to Linkverse. The remaining authors have no conflicts of interest to disclose.

References

[1] Graber ML. The incidence of diagnostic error in medicine. BMJ Qual Saf 2013;22(Suppl 2) ii21–7.

[2] Gunderson CG, Bilan VP, Holleck JL, et al. Prevalence of harmful diagnostic errors in hospitalised adults: a systematic review and meta-analysis. BMJ Qual Saf 2020;29(12):1008–18.
[3] Makary MA, Daniel M. Medical error—the third leading cause of death in the US. BMJ 2016;353:i2139.
[4] Zhao S, Wang S, Pan P, et al. Magnitude, risk factors, and factors associated with adenoma miss rate of tandem colonoscopy: a systematic review and meta-analysis. Gastroenterology 2019;156(6) 1661–74.e11.
[5] Adler J, Robertson DJ. Interval colorectal cancer after colonoscopy: exploring explanations and solutions. Off J Am Coll Gastroenterol ACG 2015;110(12):1657.
[6] Cassam Q. Diagnostic error, overconfidence and self-knowledge. Palgrave Commun 2017;3(1):1–8.
[7] Kaminski MF, Robertson DJ, Senore C, Rex DK. Optimizing the quality of colorectal cancer screening worldwide. Gastroenterology 2020;158(2):404–17.
[8] Rex DK, Schoenfeld PS, Cohen J, et al. Quality indicators for colonoscopy. Gastrointest Endosc 2015;81(1):31–53.
[9] Wolfe JM, Kosovicheva A, Wolfe B. Normal blindness: when we look but fail to see. Trends Cogn Sci 2022;26(9):809–19.
[10] Wolfe JM. Guided Search 6.0: an updated model of visual search. Psychon Bull Rev 2021;28(4):1060–92.
[11] Williams LH, Carrigan A, Auffermann W, et al. The invisible breast cancer: experience does not protect against inattentional blindness to clinically relevant findings in radiology. Psychon Bull Rev 2021;28(2):503–11.
[12] Simons DJ, Chabris CF. Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception 1999;28(9):1059–74.
[13] Casner SM, Schooler JW. Vigilance impossible: diligence, distraction, and daydreaming all lead to failures in a practical monitoring task. Conscious Cogn 2015;35:33–41.
[14] Hancock PA. On the nature of vigilance. Hum Factors 2017;59(1):35–43.
[15] Helton WS, Russell PN. Visuospatial and verbal working memory load: effects on visuospatial vigilance. Exp Brain Res 2013;224:429–36.
[16] Reiner BI, Krupinski E. The insidious problem of fatigue in medical imaging practice. J Digit Imaging 2012;25:3–6.
[17] Thomson DR, Besner D, Smilek D. A resource-control account of sustained attention: evidence from mind-wandering and vigilance paradigms. Perspect Psychol Sci 2015;10(1):82–96.
[18] Wickens CD. Multiple resources and mental workload. Hum Factors 2008;50(3):449–55.
[19] Lago MA, Jonnalagadda A, Abbey CK, et al. Under-exploration of three-dimensional images leads to search errors for small salient targets. Curr Biol 2021;31(5) 1099–106.e5.
[20] Manassi M, Whitney D. Multi-level crowding and the paradox of object recognition in clutter. Curr Biol 2018;28(3):R127–33.
[21] Võ ML-H, Aizenman AM, Wolfe JM. You think you know where you looked? You better look again. J Exp Psychol Hum Percept Perform 2016;42(10):1477–81.
[22] Williams LH, Drew T. What do we know about volumetric medical image interpretation?: a review of the basic science and medical image perception literatures. Cogn Res Princ Implic 2019;4(1):21.
[23] Howard J. Cognitive errors and diagnostic mistakes. Springer; 2019.
[24] Koehler DJ, Brenner L, Griffin D. The calibration of expert judgment: heuristics and biases beyond the laboratory. Heuristics Biases Psychol Intuitive Judgm 2002:686–715.
[25] Howard J, Howard J. Gambler's fallacy and hot hand fallacy. In: Cognitive errors and diagnostic mistakes: a case-based guide to critical thinking in medicine; 2019. p. 203–10.
[26] Nickerson RS. The production and perception of randomness. Psychol Rev 2002;109(2):330.
[27] Leufkens AM, Oijen MGH van, Vleggaar FP, Siersema PD. Factors influencing the miss rate of polyps in a back-to-back colonoscopy study. Endoscopy 2012:470–5.
[28] Wang HS, Pisegna J, Modi R, et al. Adenoma detection rate is necessary but insufficient for distinguishing high versus low endoscopist performance. Gastrointest Endosc 2013;77(1):71–8.
[29] Aniwan S, Orkoonsawat P, Viriyautsahakul V, et al. The secondary quality indicator to improve prediction of adenoma miss rate apart from adenoma detection rate. Off J Am Coll Gastroenterol ACG 2016;111(5):723.
[30] Cherubini A, East JE. Gorilla in the room: even experts can miss polyps at colonoscopy and how AI helps complex visual perception tasks. Dig Liver Dis 2023;55(2):151–3.
[31] Pashler H. Dual-task interference in simple tasks: data and theory. Psychol Bull 1994;116(2):220.
[32] Coderre S, Anderson J, Rostom A, McLaughlin K. Training the endoscopy trainer: from general principles to specific concepts. Can J Gastroenterol 2010;24(12):700–4.
[33] Jorm CM, O'Sullivan G. Laptops and smartphones in the operating theatre–how does our knowledge of vigilance, multi-tasking and anaesthetist performance help us in our approach to this new distraction? Anaesth Intensive Care 2012;40(1):71–9.
[34] Modi HN, Singh H, Darzi A, Leff DR. Multitasking and time pressure in the operating room: impact on surgeons' brain function. Ann Surg 2020;272(4):648–57.
[35] Sherbino J, Norman G. Task switching, multitasking, and errors: a psychologic perspective on the impact of interruptions. Ann Emerg Med 2021;78(3):425–8.
[36] Strobach T. The dual-task practice advantage: empirical evidence and cognitive mechanisms. Psychon Bull Rev 2020;27(1):3–14.
[37] Dawson NV, Connors AF, Speroff T, Kemka A, Shaw P, Arkes HR. Hemodynamic assessment in managing the critically iii: is physician confidence warranted? Med Decis Making 1993;13(3):258–66.
[38] Klein G, Phillips JK, Rall EL, Peluso DA. A data–frame theory of sensemaking. Expertise out of context. Psychology Press; 2007.
[39] Einhorn HJ, Hogarth RM. Confidence in judgment: persistence of the illusion of validity. Psychol Rev 1978;85(5):395.
[40] Szegedy C, Zaremba W, Sutskever I, Bruna JD, Erhan I, Fergus RGoodfellow. Intriguing, properties of neural networks. 2nd international conference on learning representations (ICLR); 2014.
[41] Hassan C, Wallace MB, Sharma P, et al. New artificial intelligence system: first validation study versus experienced endoscopists for colorectal polyp detection. Gut 2020;69(5):799–800.
[42] Rex DK, Mori Y, Sharma P, Lahr RE, Vemulapalli KC, Hassan C. Strengths and weaknesses of an artificial intelligence polyp detection program as assessed by a high-detecting endoscopist. Gastroenterology 2022;163(2) 354–8.e1.
[43] Vilmann AS, Norsk D, Svendsen MBS, et al. Computerized feedback during colonoscopy training leads to improved performance: a randomized trial. Gastrointest Endosc 2018;88(5):869–76.
[44] Behrens TEJ, Hunt LT, Woolrich MW, Rushworth MFS. Associative learning of social value. Nature 2008;456(7219):245–9.
[45] Koriat A. When are two heads better than one and why? Science 2012;336(6079):360–2.
[46] De Martino B, Bobadilla-Suarez S, Nouguchi T, Sharot T, Love BC. Social information is integrated into value and confidence judgments according to its reliability. J Neurosci 2017;37(25):6066–74.
[47] Park SA, Goïame S, O'Connor DA, Dreher J-C. Integration of individual and social information for decision-making in groups of different sizes. PLOS Biol 2017;15(6):e2001958.
[48] Sniezek JA, Van Swol LM. Trust, confidence, and expertise in a judge-advisor system. Organ Behav Hum Decis Process 2001;84(2):288–307.
[49] Larrick RP, Soll JB. Intuitions about combining opinions: misappreciation of the averaging principle. Manag Sci 2006;52(1):111–27.
[50] Bahrami B, Olsen K, Latham PE, Roepstorff A, Rees G, Frith CD. Optimally interacting minds. Science 2010;329(5995):1081–5.
[51] Bahrami B, Olsen K, Bang D, Roepstorff A, Rees G, Frith C. What failure in collective decision-making tells us about metacognition. Philos Trans R Soc B 2012;367(1594):1350–65.
[52] Bailey PE, Leon T, Ebner NC, Moustafa AA, Weidemann G. A meta-analysis of the weight of advice in decision-making. Curr Psychol 2022 [cited 2023 Jul 6];Available from:. doi:10.1007/s12144-022-03573-2.
[53] Bonaccio S, Dalal RS. Advice taking and decision-making: an integrative literature review, and implications for the organizational sciences. Organ Behav Hum Decis Process 2006;101(2):127–51.
[54] Krueger JI. Return of the ego–Self-referent information as a filter for social prediction: comment on Karniol (2003). Psychol Rev 2003;110(3):585–90.
[55] Morin O, Jacquet PO, Vaesen K, Acerbi A. Social information use and social information waste. Philos Trans R Soc B 2021;376(1828):20200052.
[56] Yaniv I. Receiving other people's advice: influence and benefit. Organ Behav Hum Decis Process 2004;93(1):1–13.
[57] Yaniv I, Kleinberger E. Advice taking in decision making: egocentric discounting and reputation formation. Organ Behav Hum Decis Process 2000;83(2):260–81.
[58] Barnett ML, Boddupalli D, Nundy S, Bates DW. Comparative accuracy of diagnosis by collective intelligence of multiple physicians vs individual physicians. JAMA Netw Open 2019;2(3):e190096.
[59] Galton F. Vox Populi. Nature 1907;75(1949):450–1.
[60] Surowiecki J. The wisdom of crowds. Knopf Doubleday Publishing Group; 2005.
[61] Green B, Chen Y. The principles and limits of algorithm-in-the-loop decision making. Proc ACM Hum-Comput Interact 2019;3(CSCW) 50:1-50:24.
[62] Leslie AM, Friedman O, German TP. Core mechanisms in 'theory of mind. Trends Cogn Sci 2004;8(12):528–33.
[63] Mercado JE, Rupp MA, Chen JYC, Barnes MJ, Barber D, Procci K. Intelligent agent transparency in human–agent teaming for multi-UxV management. Hum Factors 2016;58(3):401–15.
[64] Yang XJ, Unhelkar VV, Li K, Shah JA. Evaluating effects of user experience and system transparency on trust in automation [Internet]. In: Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction. Vienna Austria: ACM; 2017. p. 408–16. [cited 2023 Jul 12]Available from:.
[65] Cadario R, Longoni C, Morewedge CK. Understanding, explaining, and utilizing medical artificial intelligence. Nat Hum Behav 2021;5(12):1636–42.
[66] Reverberi C, Rigon T, Solari A, et al. Experimental evidence of effective human-AI collaboration in medical decision-making. Sci Rep 2022;12(1):14952.
[67] Bogert E, Schecter A, Watson RT. Humans rely more on algorithms than social influence as a task becomes more difficult. Sci Rep 2021;11(1):8028.
[68] Cabitza F, Campagner A, Ronzio L, et al. Rams, hounds and white boxes: investigating human–AI collaboration protocols in medical diagnosis. Artif Intell Med 2023;138:102506.
[69] Logg JM, Minson JA, Moore DA. Algorithm appreciation: people prefer algorithmic to human judgment. Organ Behav Hum Decis Process 2019;151:90–103.
[70] Mahmud H, Islam AKMN, Ahmed SI, Smolander K. What influences algorith-

mic decision-making? A systematic literature review on algorithm aversion. Technol Forecast Soc Change 2022;175:121390.
[71] Madhavan P, Wiegmann DA, Lacson FC. Automation failures on tasks easily performed by operators undermine trust in automated aids. Hum Factors 2006;48(2):241–56.
[72] Zonca J, Folsø A, Sciutti A. Social Influence under uncertainty in interaction with peers, robots and computers. Int J Soc Robot 2023;15(2):249–68.
[73] Bansal G, Nushi B, Kamar E, Weld DS, Lasecki WS, Horvitz E. Updates in human-AI teams: understanding and addressing the performance/compatibility tradeoff. Proc AAAI Conf Artif Intell 2019;33(01):2429–37.
[74] McGuirl JM, Sarter NB. Supporting trust calibration and the effective use of decision aids by presenting dynamic system confidence information. Hum Factors 2006;48(4):656–65.
[75] Zhang Y, Liao QV, Bellamy RKE. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making [Internet]. In: Proceedings of the 2020 conference on fairness, accountability, and transparency. Barcelona Spain: ACM; 2020. p. 295–305. [cited 2023 Jul 13]Available from:.
[76] Shah S, Park N, Chehade NEH, et al. Effect of computer-aided colonoscopy on adenoma miss rates and polyp detection: a systematic review and meta-analysis. J Gastroenterol Hepatol 2023;38(2):162–76.
[77] Hassan C, Badalamenti M, Maselli R, et al. Computer-aided detection-assisted colonoscopy: classification and relevance of false positives. Gastrointest Endosc 2020;92(4) 900–4.e4.
[78] Barua I, Vinsard DG, Jodal HC, et al. Artificial intelligence for polyp detection during colonoscopy: a systematic review and meta-analysis. Endoscopy 2021;53(3):277–84.
[79] Repici A, Badalamenti M, Maselli R, et al. Efficacy of real-time computer-aided detection of colorectal neoplasia in a randomized trial. Gastroenterology 2020;159(2) 512–20.e7.
[80] Wang P, Liu X, Berzin TM, et al. Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. Lancet Gastroenterol Hepatol 2020;5(4):343–51.
[81] Xu H, Tang RSY, Lam TYT, et al. Artificial intelligence-assisted colonoscopy for colorectal cancer screening: a multicenter randomized controlled trial. Clin Gastroenterol Hepatol 2023;21(2) 337–46.e3.
[82] Gimeno-García AZ, Hernández Negrin D, Hernández A, et al. Usefulness of a novel computer-aided detection system for colorectal neoplasia: a randomized controlled trial. Gastrointest Endosc 2023;97(3) 528–36.e1.
[83] Huang D, Shen J, Hong J, et al. Effect of artificial intelligence-aided colonoscopy for adenoma and polyp detection: a meta-analysis of randomized clinical trials. Int J Colorectal Dis 2022;37(3):495–506.
[84] Wallace MB, Sharma P, Bhandari P, et al. Impact of artificial intelligence on miss rate of colorectal neoplasia. Gastroenterology 2022;163(1) 295–304.e5.
[85] Lu Z, Zhang L, Yao L, et al. Assessment of the role of artificial intelligence in the association between time of day and colonoscopy quality. JAMA Netw Open 2023;6(1):e2253840.
[86] Gurudu SR, Ratuapli SK, Leighton JA, Heigh RI, Crowell MD. Adenoma detection rate is not influenced by the timing of colonoscopy when performed in half-day blocks. Off J Am Coll Gastroenterol ACG 2011;106(8):1466.
[87] Lee A, Iskander JM, Gupta N, et al. Queue position in the endoscopic schedule impacts effectiveness of colonoscopy. Off J Am Coll Gastroenterol ACG 2011;106(8):1457.
[88] Lee TJW, Rees CJ, Blanks RG, et al. Colonoscopic factors associated with adenoma detection in a national colorectal cancer screening program. Endoscopy 2014:203–11.
[89] Ahmad A, Wilson A, Haycock A, et al. Evaluation of a real-time computer-aided polyp detection system during screening colonoscopy: AI-DETECT study. Endoscopy 2022:313–19.
[90] Ladabaum U, Shepard J, Weng Y, Desai M, Singer SJ, Mannalithara A. Computer-aided detection of polyps does not improve colonoscopist performance in a pragmatic implementation trial. Gastroenterology 2023;164(3) 481–3.e6.
[91] Levy I, Bruckmayer L, Klang E, Ben-Horin S, Kopylov U. Artificial intelligence-aided colonoscopy does not increase adenoma detection rate in routine clinical practice. Am J Gastroenterol 2022;117(11):1871–3.
[92] Berzin TM, Glissen Brown J. Navigating the "Trough of Disillusionment" for CADe polyp detection: what can we learn about negative ai trials and the physician-AI hybrid? Off J Am Coll Gastroenterol ACG 2022. doi:10.14309/ajg.0000000000002286.
[93] Cabitza F, Campagner A, Angius R, Natali C, Reverberi C. AI shall have no dominion: on how to measure technology dominance in AI-supported human decision-making [Internet]. In: Proceedings of the 2023 CHI conference on human factors in computing systems. New York, NY, USA: Association for Computing Machinery; 2023. p. 1–20. [cited 2023 Apr 20]Available from:.
[94] Barua I, Wieszczy P, Kudo S, et al. Real-time artificial intelligence–based optical diagnosis of neoplastic polyps during colonoscopy. NEJM Evid 2022;1(6):EVIDoa2200003.
[95] Hassan C, Sharma P, Mori Y, et al. Comparative performance of artificial intelligence optical diagnosis systems for leaving in situ colorectal polyps. Gastroenterology 2023;164(3):467–9 .e4.
[96] Rondonotti E, Hassan C, Tamanini G, et al. Artificial intelligence-assisted optical diagnosis for the resect-and-discard strategy in clinical practice: the Artificial intelligence BLI Characterization (ABC) study. Endoscopy 2022:14–22.
[97] Druce J., Niehaus J., Moody V., Jensen D., Littman M.L. Brittle AI, Causal Confusion, and Bad Mental Models: Challenges and Successes in the XAI Program [Internet]. 2021 [cited 2023 Jul 27];Available from: http://arxiv.org/abs/2106.05506
[98] Sloman SA, Lagnado D. Causality in thought. Annu Rev Psychol 2015;66(1):223–47.
[99] Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. JAMA 2016;316(22):2353–4.
[100] Slovic P. Perception of risk. Science 1987;236(4799):280–5.
[101] Deliwala SS, Hamid K, Barbarawi M, et al. Artificial intelligence (AI) real-time detection vs. routine colonoscopy for colorectal neoplasia: a meta-analysis and trial sequential analysis. Int J Colorectal Dis 2021;36(11):2291–303.
[102] Hassan C, Spadaccini M, Iannone A, et al. Performance of artificial intelligence in colonoscopy for adenoma and polyp detection: a systematic review and meta-analysis. Gastrointest Endosc 2021;93(1) 77–85.e6.
[103] Mori Y, Wang P, Løberg M, et al. Impact of artificial intelligence on colonoscopy surveillance after polyp removal: a pooled analysis of randomized trials. Clin Gastroenterol Hepatol 2023;21(4) 949–59.e2.
[104] Spadaccini M, Iannone A, Maselli R, et al. Computer-aided detection versus advanced imaging for detection of colorectal neoplasia: a systematic review and network meta-analysis. Lancet Gastroenterol Hepatol 2021;6(10):793–802.
[105] Zhang Y, Zhang X, Wu Q, Gu C, Wang Z. Artificial intelligence-aided colonoscopy for polyp detection: a systematic review and meta-analysis of randomized clinical trials. J Laparoendosc Adv Surg Tech A 2021;31(10):1143–9.
[106] Glissen Brown JR, Mansour NM, Wang P, et al. Deep learning computer-aided polyp detection reduces adenoma miss rate: a United States Multi-center Randomized Tandem Colonoscopy Study (CADeT-CS Trial). Clin Gastroenterol Hepatol 2022;20(7) 1499–507.e4.
[107] Gong D, Wu L, Zhang J, et al. Detection of colorectal adenomas with a real–time computer-aided system (ENDOANGEL): a randomised controlled study. Lancet Gastroenterol Hepatol 2020;5(4):352–61.
[108] Hüneburg R, Bucksch K, Schmeißer F, et al. Real-time use of artificial intelligence (CADEYE) in colorectal cancer surveillance of patients with Lynch syndrome-A randomized controlled pilot trial (CADLY). United. Eur Gastroenterol J 2023;11(1):60–8.
[109] Kamba S, Tamai N, Saitoh I, et al. Reducing adenoma miss rate of colonoscopy assisted by artificial intelligence: a multicenter randomized controlled trial. J Gastroenterol 2021;56(8):746–57.
[110] Karsenti D, Tharsis G, Perrot B, et al. Effect of real-time computer-aided detection of colorectal adenoma in routine colonoscopy (COLO-GENIUS): a single-centre randomised controlled trial. Lancet Gastroenterol Hepatol 2023;8(8):726–34.
[111] Liu W-N, Zhang Y-Y, Bian X-Q, et al. Study on detection rate of polyps and adenomas in artificial-intelligence-aided colonoscopy. Saudi J Gastroenterol 2020;26(1):13.
[112] Luo Y, Zhang Y, Liu M, et al. Artificial intelligence-assisted colonoscopy for detection of colon polyps: a prospective, randomized cohort study. J Gastrointest Surg 2021;25(8):2011–18.
[113] Lux TJ, Banck M, Saßmannshausen Z, et al. Pilot study of a new freely available computer-aided polyp detection system in clinical practice. Int J Colorectal Dis 2022;37(6):1349–54.
[114] Nakashima H, Kitazawa N, Fukuyama C, et al. Clinical evaluation of computer-aided colorectal neoplasia detection using a novel endoscopic artificial intelligence: a single-center randomized controlled trial. Digestion 2023;104(3):193–201.
[115] Quan SY, Wei MT, Lee J, et al. Clinical evaluation of a real-time artificial intelligence-based polyp detection system: a US multi-center pilot study. Sci Rep 2022;12(1):6598.
[116] Repici A, Spadaccini M, Antonelli G, et al. Artificial intelligence and colonoscopy experience: lessons from two randomised trials. Gut 2022;71(4):757–65.
[117] Shaukat A, Lichtenstein DR, Somers SC, et al. Computer-aided detection improves adenomas per colonoscopy for screening and surveillance colonoscopy: a randomized trial. Gastroenterology 2022;163(3):732–41.
[118] Su J-R, Li Z, Shao X-J, et al. Impact of a real-time automatic quality control system on colorectal polyp and adenoma detection: a prospective randomized controlled study (with videos). Gastrointest Endosc 2020;91(2) 415–24.e4.
[119] Wang P, Liu P, Glissen Brown JR, et al. Lower adenoma miss rate of computer-aided detection-assisted colonoscopy vs routine white-light colonoscopy in a prospective tandem study. Gastroenterology 2020;159(4) 1252–61.e5.
[120] Wei MT, Shankar U, Parvin R, et al. Evaluation of computer-aided detection during colonoscopy in the community (AI-SEE): a multicenter randomized clinical trial. Am J Gastroenterol 2023.
[121] Xu L, He X, Zhou J, et al. Artificial intelligence-assisted colonoscopy: a prospective, multicenter, randomized controlled trial of polyp detection. Cancer Med 2021;10(20):7184–93.
[122] Yamaguchi D, Shimoda R, Miyahara K, et al. Impact of an artificial intelligence-aided endoscopic diagnosis system on improving endoscopy quality for trainees in colonoscopy: prospective, randomized, multicenter study. Dig Endosc 2023.
[123] Yao L, Zhang L, Liu J, et al. Effect of an artificial intelligence-based quality improvement system on efficacy of a computer-aided detection system in colonoscopy: a four-group parallel study. Endoscopy 2022;54(8):757–68.