Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy

van de Sande, Davy; Sleijfer, Stefan; Buijsman, Stefan N.R.; Gommers, Diederik; van Genderen, Michel E.

DOI
10.1016/S2468-1253(25)00288-2

Publication date
2025

Document Version
Final published version

Published in
The lancet. Gastroenterology & hepatology

impact on serrated polyp detection despite improving adenoma detection. Even more concerning is the effect on technique and visual search patterns. Troya and colleagues[4] show that AI assistance significantly reduces eye travel distance during mucosal inspection and increases misinterpretation of normal mucosa as polyps. This implies endoscopists using AI systems wait for algorithmic cues rather than maintain systematic visual searching (basically looking for AI markings and not for polyps). Current AI devices do not address quality metrics and do not alert the endoscopist to pitfalls such as potential mucosal underexposure or insufficient withdrawal time. Such changes risk missing lesions not captured by AI systems and might prevent trainees from developing optimal visual pattern characteristic of high-performing endoscopists.

Training programmes must urgently implement pragmatic mitigations. Suggested strategies include periodic AI-off sessions in which fellows develop endoscopic skills without technological assistance, structured competency assessments, and specific modules on serrated polyp detection and early neoplasia recognition. Longitudinal studies tracking performance of young endoscopists are essential to understand long-term effects of AI exposure.[5] Professional societies (eg, the European Society of Gastrointestinal Endoscopy) are developing guidelines for endoscopy training in AI-enhanced environments; however, these are constructed on scarce evidence.

As we embrace the wide potential of AI, we must ensure technological enhancement does not compromise the comprehensive diagnostic and cognitive skills and fundamental techniques defining expert practice.

UK reports research support from AbbVie, BMS, Eli Lilly, Jannsen, Medtronic, and Takeda, and speaker and advisory fees from AbbVie, BMS, Celltrion, Eli Lilly, Medtronic, Pfizer, and Takeda. AL declares no competing interests.

*Asaf Levartovsky, Uri Kopylov
alavra@gmail.com

Department of Gastroenterology, Sheba Medical Center, Ramat-Gan, Israel (AL, UK); School of Medicine, Faculty of Medical and Health Sciences, Tel-Aviv University, Tel-Aviv, Israel (AL, UK)

1 Budzyń K, Romańczyk M, Kitala D, et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. *Lancet Gastroenterol Hepatol* 2025; **10:** 896–903.
2 Lyell D, Coiera E. Automation bias and verification complexity: a systematic review. *J Am Med Inform Assoc* 2017; **24:** 423–31.
3 van Toledo DEFWM, IJspeert JEG, Bossuyt PMM, et al. Serrated polyp detection and risk of interval post-colonoscopy colorectal cancer: a population-based study. *Lancet Gastroenterol Hepatol* 2022; **7:** 747–54.
4 Troya J, Fitting D, Brand M, et al. The influence of computer-aided polyp detection systems on reaction time for polyp detection and eye gaze. *Endoscopy* 2022; **54:** 1009–14.
5 Kang AJ, Rodrigues T, Patel RV, Keswani RN. Impact of artificial intelligence on gastroenterology trainee education. *Gastrointest Endosc Clin N Am* 2025; **35:** 457–67.

The multicentre observational study by Krzysztof Budzyń and colleagues[1] reports that after routine exposure to artificial intelligence (AI)-assisted polyp detection, endoscopists' performance during non-AI colonoscopy fell by 6·0%. Some will interpret these findings as evidence that AI exposure affects doctors' interpretation abilities and as a reason to retreat from assistance, an intuitive yet incomplete reading.

Two distinct phenomena are relevant here—automation bias and skill reallocation. Automation bias is an acute bedside judgement failure, where clinicians defer to an AI suggestion even when it is incorrect.[2] Skill reallocation is longitudinal: when a tool reliably takes over a task or part of a task, clinicians practice that part less and redirect attention to other tasks. For example, with consistent AI cueing, endoscopists might spend less time on frame-by-frame vigilance and more on complete mucosal exposure (including blind spots), differentiation of clinically relevant lesions, and appropriate next steps. Such reallocation is not inherently harmful. It becomes unsafe only when institutions do not preserve the residual human capabilities required when automation is absent, degraded, or misleading.

Seen this way, Budzyń and colleagues highlight a governance and ethics challenge—not a reason to retreat from AI. Where robust evidence shows that AI improves outcomes or efficiency, it has to be decided which tasks in a care pathway can be shifted from health professionals to AI. Meaningful human control requires teams to recognise—and act decisively—when automation misleads. Therefore, health-care organisations need dual-track competence: baseline vigilance without AI, and calibrated judgement with it. This entails auditing with-AI and without-AI performance as separate indicators, scheduling deliberate AI-off practice (ie, periods of routine work or simulation with the AI system disabled to maintain unaided detection skills), and designing transparent interfaces that expose uncertainty and show the visual basis of cues. Also, AI literacy—including EU AI Act duties—should be embedded in training.[3]

Budzyń and colleagues show that skill reallocation is already happening when working with AI.[1] The ethical response is not alarmism about deskilling but purposeful clinical redesign. Skill reallocation, properly managed, is part of building clinical trust in AI—much as with calculators and autopilot.

We declare no competing interests.

*Davy van de Sande, Stefan Sleijfer, Stefan N R Buijsman, Diederik Gommers, *Michel E van Genderen
m.vangenderen@erasmusmc.nl

Erasmus MC University Medical Center, Department of Adult Intensive Care, Rotterdam 3015, Netherlands (DvdS, DG, MEvG); Erasmus MC Datahub, University Center Rotterdam, Rotterdam, Netherlands (DvdS, DG, MEvG); Department of Medical Oncology, Erasmus MC Cancer Institute, Erasmus University Medical Center, Rotterdam, Netherlands (SS); Executive Board, Erasmus MC, Rotterdam, Netherlands (SS); Delft University of Technology, Faculty of Technology, Policy and Management, Delft, Netherlands (SNRB)

1 Budzyń K, Romańczyk M, Kitala D, et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. *Lancet Gastroenterol Hepatol* 2025; **10:** 896–903.
2 Tikhomirov L, Semmler C, McCradden M, Searston R, Ghassemi M, Oakden-Rayner L. Medical artificial intelligence for clinicians: the lost cognitive perspective. *Lancet Digit Health* 2024; **6:** e589–94.
3 European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). European Union, 2024.

Authors' reply

We thank Kyle Lam, Asaf Levartovsky and colleagues, and Davy van de Sande and colleagues for their insightful comments highlighting the importance of human–artificial intelligence (AI) interaction.

Interaction with AI is a clinically relevant yet controversial issue. Our study showed that continuous exposure to AI during colonoscopy was associated with a reduction in adenoma detection rate (ADR) once AI was withdrawn.[1] Similar effects have been observed in other fields, such as aviation, in which excessive automation has diminished pilots' manual skills.[2] Another research group suggested that exposure to AI might reduce visual engagement during colonoscopy, indicating a possible automation complacency effect.[3] Psychological studies further support the notion that over-reliance on AI can weaken critical thinking, potentially diminishing long-term professional competence. Interventional research in behavioural science suggests that cognitive-forcing or psychological nudging techniques might help restore critical thinking and optimise AI use.[4] Despite growing interest, few studies have clinically validated these phenomena using patient-relevant outcomes such as ADR. Therefore, we are pleased that our hypothesis-generating work has stimulated constructive discussion in this emerging field.

Lam emphasises the potential influence of AI on less experienced endoscopists. We agree that this is an important consideration—particularly for medical education and training. However, current evidence, including ours, remains scarce. Well-designed clinical trials are needed to clarify this issue.

As noted by Levartovsky and colleagues, sessile serrated lesions differ substantially from conventional adenomas in appearance. We concur that the potential deskilling effect could be more pronounced when sessile serrated lesions are included, as their detection requires higher cognitive effort. Again, this hypothesis also warrants testing in clinical trials.

Another important point raised concerns whether withholding AI might disadvantage patients and physicians in light of our findings. Although several randomised controlled trials have shown that AI improves ADR and other quality indicators, many observational studies have reported no such benefit.[5] As we discussed in our paper, control groups in randomised trials might be indirectly influenced by a deskilling effect, potentially exaggerating the apparent benefit of AI. However, our study was designed to generate hypotheses rather than to provide definitive conclusions. The next step is to conduct rigorously designed trials to objectively assess AI's effect on human performance, with careful attention to control conditions and potential time effects on skill acquisition.

In conclusion, our findings of reduced performance following AI withdrawal have sparked important discussion. The field now needs high-quality clinical studies to determine whether this effect is genuine, and if confirmed, to develop strategies to mitigate it. Colonoscopy offers an ideal model for such work, given the availability of robust, patient-oriented outcome measures.

YM has received consultancy fees, speaker's fees, and a device loan from Olympus, royalty fees from Cybernet System, and financial support from the European Commission (Horizon Europe 101057099) and the Japan Society for the Promotion of Science (22H03357). KB and MR declare no competing interests.

*Krzysztof Budzyń, *Marcin Romańczyk, Yuichi Mori*
marcin.romanczyk@htcentrum.pl

Department of Gastroenterology, Faculty of Medicine, Academy of Silesia, Katowice, Poland (KB, MR); Endoterapia, H-T Centrum Medyczne, 43-100 Tychy, Poland (KB, MR); Clinical Effectiveness Research Group, University of Oslo and Oslo University Hospital, Oslo, Norway (YM); Gastroenterology Section, Department of Transplantation Medicine, Oslo University Hospital, Oslo, Norway (YM); Digestive Disease Center, Showa University Northern Yokohama Hospital, Yokohama, Japan (YM)

1 Budzyń K, Romańczyk M, Kitala D, et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. *Lancet Gastroenterol Hepatol* 2025; **10:** 896–903.
2 Haslbeck A, Hoermann H-J. Flying the needles: human performance consequences of automation in aviation. *Hum Factors* 2016; **58:** 533–45.
3 Troya J, Fitting D, Brand M, et al. The influence of computer-aided polyp detection systems on reaction time for polyp detection and eye gaze. *Endoscopy* 2022; **54:** 1009–14.
4 Kunar MA, Watson DG. Framing the fallibility of computer-aided detection aids cancer detection. *Cogn Res Princ Implic* 2023; **8:** 30.
5 Patel HK, Mori Y, Hassan C, et al. Lack of effectiveness of computer aided detection for colorectal neoplasia: a systematic review and meta-analysis of nonrandomized studies. *Clin Gastroenterol Hepatol* 2024; **22:** 971–80.