caisa Det Nationale Center for AI i Samfundet
The National Center for AI in Society

Adoption of AI Diagnostics in General Practice: Promises and Challenges

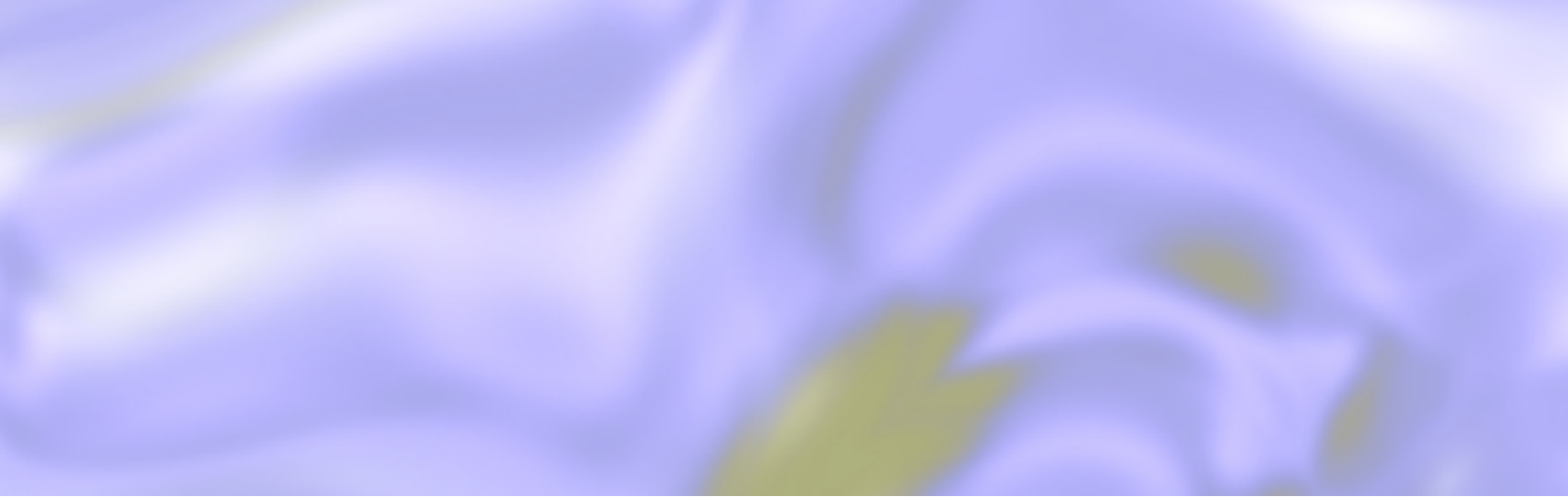
CAISA Research Brief

Authors
Ullrich, Hannes; Huang, Shan; Schmacker, Renke

Editors
Søgaard, Anders; Feldt, Johannes N.

- Publication date: 24 June 2026
- Published by: CAISA, The National Center for AI in Society, Copenhagen and Aalborg, Denmark

- ISSN: 2795-0646
- Document version: Publisher's PDF, registered version
- Citation for published version (APA): Ullrich, H., Huang, S., & Schmacker, R. (2026). Adoption of AI Diagnostics in General Practice: Promises and Challenges. *CAISA - Brief*. www.caisa.dk/en/forskning/adoption-of-ai-diagnostics-in-general-practice-promises-and-challenges



Adoption of AI Diagnostics in General Practice: Promises and Challenges

By Hannes Ullrich, Shan Huang, and Renke Schmacker

Summary

Artificial intelligence (AI) is entering general practice (GP) through documentation tools, digital assistants, and diagnostic support systems. Even though AI is a powerful prediction technology with impressive potential for health care applications, the key question is how AI tools will be designed and whether clinicians will be able to use them in ways that improve care. New evidence from a countrywide survey experiment in Denmark shows potential challenges of AI adoption in general practice. Around one-third of general practitioners are not comfortable following AI predictions even if they contain meaningful information on disease risk extracted from data on patients' medical histories. Those who do use the AI-generated information use this information differently than information generated by a well-known diagnostic test that is comparable in quality to the AI tool. These findings point to the importance of how diagnostic information is presented and explained for achieving the potential of new prediction technologies. Therefore, the challenge is not the development of the technology alone, but the implementation so that it is adopted in practice and generates value for physicians and patients. To achieve this, efforts to advance implementations should involve empirical evaluations of how AI-generated information is communicated to clinicians, accounting for the diagnostic tools and routines already in place, for organizational and behavioral components of clinical practice, and how their interaction affects treatment outcomes.

Main Points

1. AI is entering general practice but, due to its interaction with organizational and behavioral factors, its adoption and ensuing consequences are not yet well understood.

2. In a country-wide experiment in Denmark, general practitioners changed their assessments of UTI risk by 18.5 percentage points after a familiar dipstick test, but by only 10.9 percentage points after an AI-based diagnostic test – a 41% smaller response.

3. Uptake was uneven: roughly two-thirds of physicians used the AI tool to varying extents, while about one-third barely used it at all.

4. Use of the AI tool was more limited among general practitioners with concerns about whether it usefully extracted diagnostic information from patient data, and in clinics with generally lower technology affinity.

5. To support beneficial AI adoption in general practice, practice-based evaluation of how clinicians use AI in practice should be combined with effective training during implementation.

Background and Purpose

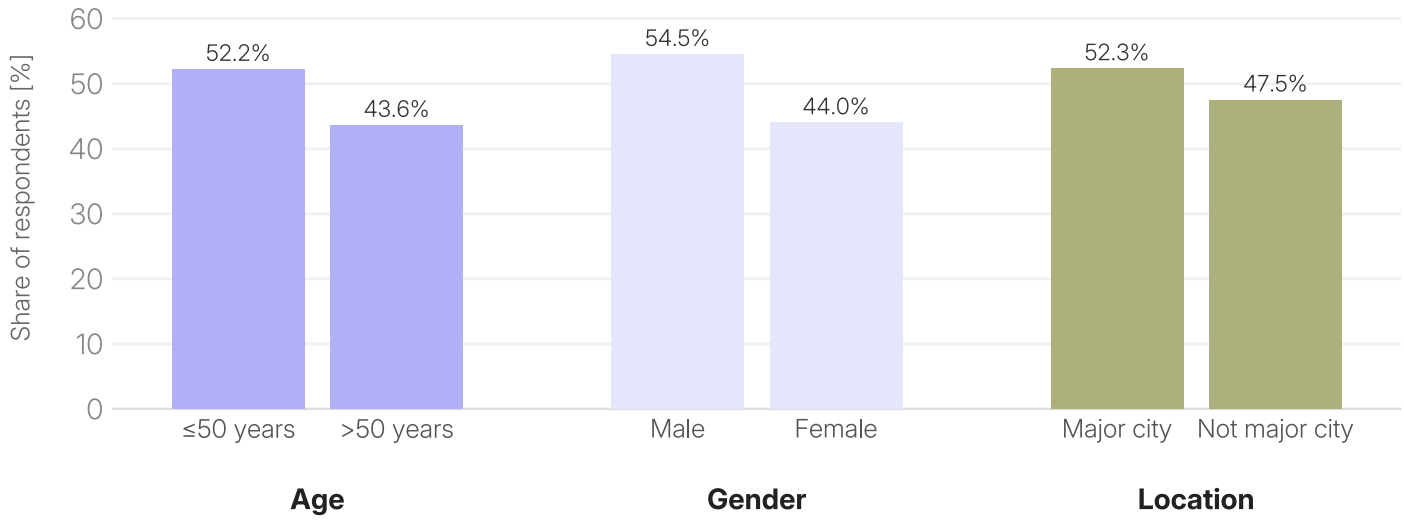
Artificial intelligence is increasingly supporting general practice by easing workload and improving access to treatment-relevant information, where GPs routinely manage multi-morbidity, medically unexplained symptoms, shared decision-making, and coordination across sectors. Vendors have introduced generative AI tools that can retrieve patient information, draft notes, reduce administrative burdens, and assist clinical reasoning that already go beyond what GPs perceived to be acceptable and useful in recent years (Jørgensen et al., 2025). For diagnostic classification tasks, AI can combine information from symptoms, clinical histories, patient characteristics, and test results more quickly and systematically than humans. But improved information alone does not guarantee better care. In general practice, clinicians make decisions under time pressure, with incomplete information, and in long-term relationships with patients. Physicians typically hold important contextual knowledge about individual patients that is difficult to encode in AI systems, and that ideally needs to be combined with the information an AI can deliver. Whether AI improves care therefore depends not only on technical accuracy, but also on whether physicians trust a tool, understand what it contributes, and use it appropriately in combination with their existing situation-specific knowledge.

This brief discusses the challenges for AI adoption and use in general practice through the findings of a new study (Huang et al., 2026).[1] Using a survey experiment, the study zooms in on a diagnostic classification task and investigates how Danish GPs might respond to AI-based diagnostic support in a clinically important setting: urinary tract infections (UTI). Studying AI in this context is useful because patients with UTI symptoms are common in primary care and diagnostic uncertainty is high at first contact (Holm et al., 2020; Saust et al., 2022). Inefficient use of antibiotics is one important consequence of such diagnostic uncertainty, complicating the political goal of reducing antibiotic use to curb antibiotic resistance (Huang and Ullrich, 2026). The diagnosis of UTI represents a typical classification task in general practice where AI has been shown to have the potential to reduce diagnostic uncertainty and improve treatment (Kanjilal et al., 2020; Huang et al., 2022; Ribers and Ullrich 2024).

AI adoption is already a present issue rather than a distant one: a 2024 UK survey found that one in five GPs reported using generative AI tools in clinical practice for documentation and diagnostic support (Blease et al., 2024). In the study presented in this brief, in 2025, 48% of Danish GPs stated they use generative AI. Figure 1 shows that generative AI use in clinical practice is reported more often by younger and male physicians, and by clinics located in urban areas. These findings are consistent with the unequal adoption of AI that has been documented across occupations in Denmark (Humlum and Vestergaard, 2025).

In addition, large shares of participating GPs state that they believe AI will be helpful for summarizing patient information (82%), administrative support (67%), and diagnostic support (49%). These observations call for systematic evaluation of how AI reshapes decision-making in general practice, and how AI tools need to be designed and implemented to lead to improved care (Katonai et al., 2025).

Figure 1: Respondents using generative AI in general practice, key characteristics.



[1] The study can be accessed at: https://www.diw.de/documents/publikationen/73/diw_01.c.1010776.de/dp2168.pdf

What the Study Did

The study reports results of a nationwide survey experiment, conducted in Denmark in 2025, with 372 general practitioners which represent over one fifth of all general practice clinics.[2] Physicians evaluated patient cases presenting with possible UTI and made hypothetical diagnostic and prescribing decisions before and after receiving additional diagnostic information.[3]

In the experiment, physicians were randomly assigned to receive one of two pieces of diagnostic information, indicating whether a bacterial infection is likely to be present or not. The first was a urine dipstick test, which is well-known to physicians. The dipstick delivers two main measures that help to diagnose a UTI: the presence of nitrites, that bacteria produce from naturally occurring nitrates, and leukocyte esterase, an enzyme produced by white blood cells that appear with an active infection. The second was a new AI support tool that predicts laboratory-confirmed bacterial infection using patient data *and* the dipstick result. The AI was presented as a prediction tool, not a recommendation or conclusive diagnostic, together with references to leading medical publications that document the ability of such an AI tool to extract diagnostic information from patient data that are new and valuable for diagnosis.

Crucially, in the experiment, the two diagnostics were designed to have the same statistical properties. Because they are equally accurate, a physician who responds only to the substance of the result and not to where it came from should treat them the same way. Thus, any difference in reactions reveals how the source of the information, rather than its quality, shapes diagnostic assessments and treatment decisions.

Main Findings on the Adoption of an AI Tool

The core finding is straightforward: physicians changed their diagnostic assessments less when the information came from the AI tool than when it came from the dipstick, even though the two were equally informative. The AI changed physicians' assessments of UTI probabilities by 10.9 percentage points. After seeing the dipstick test result, GPs changed their assessment by 18.5 percentage points. This difference means GPs were 41% less responsive to the AI tool compared to the well-known dipstick tool.

A further finding, with consequences for AI design and communication, concerns how physicians interpret the direction of a diagnostic signal. Doctors are used to reading medical tests on simplified scales, or even as yes-or-no results, that are better at confirming a "no" than a "yes" (or the other way around): the dipstick functions as a "rule-out" test, where detecting no white blood cells rules out an infection quite well. The AI tool in our experiment also gives a simple yes-or-no result, which produces more false positives than false negatives (just like the dipstick test), a property that was prominently communicated to physicians. Such asymmetry is often deliberate to avoid missing a diagnosis, for example resulting in an untreated infection. The study finds that physicians did not account for this asymmetry of the new AI tool: they treated the AI signal as if a positive and a negative result were equally accurate, when in fact they were not.

Who Adopts AI?

The average effects mask two distinct groups of GPs. About one-third (34%) were "non-adopters" of the AI tool – physicians who barely revise their assessments at all. The remaining two-thirds, the "adopters", engaged with the AI information. Adoption of the AI tool was thus low compared to the dipstick and unevenly distributed across GPs.

Why were some physicians reluctant to use the AI tool? When asked, adopters tended to emphasize that the AI brought together complementary data and were persuaded by its accuracy. Non-adopters, by contrast, questioned whether the tool performed well and whether its workings were transparent. Notably, their reservations centered on tool performance and transparency rather than on fundamental concerns about algorithmic bias, privacy, or accountability. Physicians who believed the tool meaningfully drew on relevant patient data were most likely to use the AI.[4]

Furthermore, non-adopters were more likely to work in clinics that already make less use of other digital and

[2] Even though GPs self-selected to participate, all GP clinics were invited, and the participating GPs closely resemble the population of GP clinics on a large set of characteristics.

[3] Ideally, evaluations would be performed in clinical practice with real stakes but, for multiple reasons, such interventions are difficult and have their own limitations. While the study design allows measurement of important behavioral mechanisms, behavior under real clinical conditions will be more complex.

[4] These results add to findings from Danish GP that trust in AI depends on factors including the preservation of professional autonomy, meaningful integration into workflows, transparency regarding outputs, maintenance of the GP-patient relationship, and perceived usefulness in daily practice (Jørgensen et al., 2026). Thus, trust is not merely a property of technology itself but also of its organizational context and implementation strategy.

diagnostic technologies in their clinical practice. Non-adopters did not differ in age, gender, geographic location, their perceived and observed intensity of prescribing antibiotics, or reported concerns about antibiotic resistance. This suggests that AI may reinforce existing differences in technology use across clinics, but without necessarily interacting with fundamental differences in practice styles and preferences regarding the use of antibiotics.

When more Information Makes Decisions Harder

A second finding of the study concerns how clinicians handle multiple pieces of information. When physicians were shown both the AI result and the dipstick result, they responded to the dipstick result even though it was included in the AI result, and thus redundant. This is more than a technical curiosity. It suggests that simply giving clinicians more information does not necessarily improve diagnosis. If several indicators partly reflect the same evidence, people appear to give that evidence too much weight.

This human behavior has direct implications for implementation. In some settings, AI may work better when it brings several diagnostic inputs together into one recommendation rather than presenting clinicians with several separate results. Alternatively, AI-generated signals could be designed to ensure it makes available only diagnostic information that is otherwise not accessible to physicians. In practice, decision support needs to be designed around how expert users interpret new information, and these users need to be trained accordingly. In this process, trade-offs between optimal information-processing, transparency, and human agency become important.

From Prediction to Practice

The broader lesson is that AI in healthcare needs to be understood as a tool used by clinicians. Research on this human element is sparse but an experiment with radiologists in the US finds that AI assistance with higher average diagnostic accuracy than human radiologists did not improve average performance due to human biases in the use of AI (Agarwal et al., 2025). Best outcomes were achieved with delegation, where decisions were partly made by radiologists without AI assistance and partly based on the AI tool alone. Another study in Nigeria finds that a generative AI tool improved clinical processes but not treatment outcomes (Abaluck et al., 2026).

Evidence from Danish primary care points in a similar direction: benefits may be most easily achieved by delegating between, rather than combining, algorithmic tools and clinician judgment (Ribers and Ullrich, 2024). These findings pose important challenges to the need for human agency and value of clinician expertise in AI implementations, in particular where physicians hold important context-specific information the AI does not have access to, such as patients' symptom-reporting, behavioral and social factors, or clinical assessment at the point-of-care (Ribers and Ullrich, 2023).

General practice is a particularly important setting for careful AI implementation and evaluation. GPs communicate with patients, interpret symptoms in context, manage uncertainty, and make treatment decisions shaped by clinical experience as well as evidence. Diagnosis resembles a classification task that relies on a bundle of information. Diagnostic and treatment decisions will depend on whether physicians adopt an AI that brings these pieces together for the physician or one that adds additional pieces of information that physicians bring into their evidence base.

Therefore, evaluation should not stop at validating technical accuracy and robustness, but ask how AI changes clinicians' judgments, workflow, documentation practices, referrals, testing behavior, and patient outcomes (see also "Show Us the Evidence", 2026).

Potential Consequences for Care Quality

Even though the study is only a small piece of the puzzle of AI implementation in general practice, it highlights patterns that are likely to matter for the quality of care.

In the study, physicians' underreaction to negative AI signals made them more likely to prescribe antibiotics. Similarly, when physicians received the redundant dipstick result in addition to the AI tool, they were more likely to prescribe an antibiotic when the presence of a bacterial infection was predicted, compared to physicians who did not see the redundant piece of information. These (unintended) effects matter in a setting where antibiotic use drives resistance, where, every year, an estimated 1.27 million deaths are directly attributable to antibiotic resistance.

AI will likely affect prescribing behavior, but whether the effect is beneficial will depend on design, training, and implementation rather than on predictive performance alone.

Relevant Policy Implications

With AI entering clinical practice, patients in Denmark have been found to be receptive if GPs remain responsible for decisions, AI supports rather than replaces humans, trust

in the GP-patient relationship is preserved, and data governance is transparent (Mikkelsen et al., 2023). With patients and the GP in the loop, a key policy challenge is how to evaluate the effects of AI on care delivery and to guide effective implementation in practice. Four principles are important:

First, **AI should be assessed as a support tool for expert decision-makers** who must reason under uncertainty. Trials and pilots should measure not only accuracy, but also how clinicians use the tool, how it affects behavior, and how it changes treatment outcomes. Procurement and pilot strategies could include evaluation plans that treat human-AI interaction as a central factor.

Second, **training is crucial**. Research has shown that the intention to adopt AI is high among Danish GPs (Jørgensen et al., 2026). But current research, including the new study, shows that uptake depends on whether physicians understand and believe that the tool uses relevant data well. Training and practical experience can help physicians identify when pieces of information might be redundant, when diagnostic results should be questioned, and how AI fits with professional judgment.

Third, **attention should be paid to variation in AI adoption**. If clinics with fewer diagnostic resources and generally more limited technology use are slower to adopt new tools, AI may reinforce gaps in technology use among physicians.

Fourth, **further prospective implementation studies and real-world evaluations**, including co-design with GPs as end-users, **are needed** before large-scale deployment is complete. Such efforts can provide opportunities for training, workflow redesign, integration in electronic health record systems, and evaluation of (unintended) consequences. AI implementations in administrative workflows, like documentation support, consultation summarization, inbox management, and coding, may represent a valuable, low-risk entry point to evaluate and foster AI adoption, trust, and learning.

Denmark has a well-organized and digitized primary care sector, a uniquely strong data infrastructure, and great conditions for interdisciplinary evaluation and implementation research, but access to electronic health record data and clinical software, that acts as the interface between AI and physicians, remains a barrier. If stakeholders can build on this position and overcome remaining challenges, Danish society will have the opportunity to shape how AI transforms clinical practice based on evidence reflecting the needs of general practice in Denmark.

About the authors

Hannes Ullrich is Associate Professor of Health Economics at the Department of Economics, University of Copenhagen, and Deputy Head of the Department Firms and Markets at DIW Berlin. He held a fellowship at CAISA, The National Center for AI in Society in February and March 2026.

Shan Huang is a Postdoctoral Researcher at the Department of Economics, Stockholm School of Economics.

Renke Schmacker is a Postdoctoral Researcher at the Department Public Economics at DIW Berlin and at the Research Unit Market Behavior at WZB.

About CAISA

The National Center for Artificial Intelligence in Society (CAISA) is a national consortium that gathers researchers from the University of Copenhagen, Aalborg University, Aarhus University, the IT University of Copenhagen and the Technical University of Denmark in close collaboration with the Pioneer Centre for Artificial Intelligence (P1).

As Denmark's independent research center for artificial intelligence in society, CAISA centers the citizen. We carry out groundbreaking interdisciplinary research, and we deliver overview of the most recent scientific breakthroughs. Based on new and interdisciplinary research, we advise decision-makers in the public and private sectors on how they may develop and use artificial intelligence practically, such that it contributes to growth, supports democracy and strengthens digital autonomy.

About CAISA's briefs

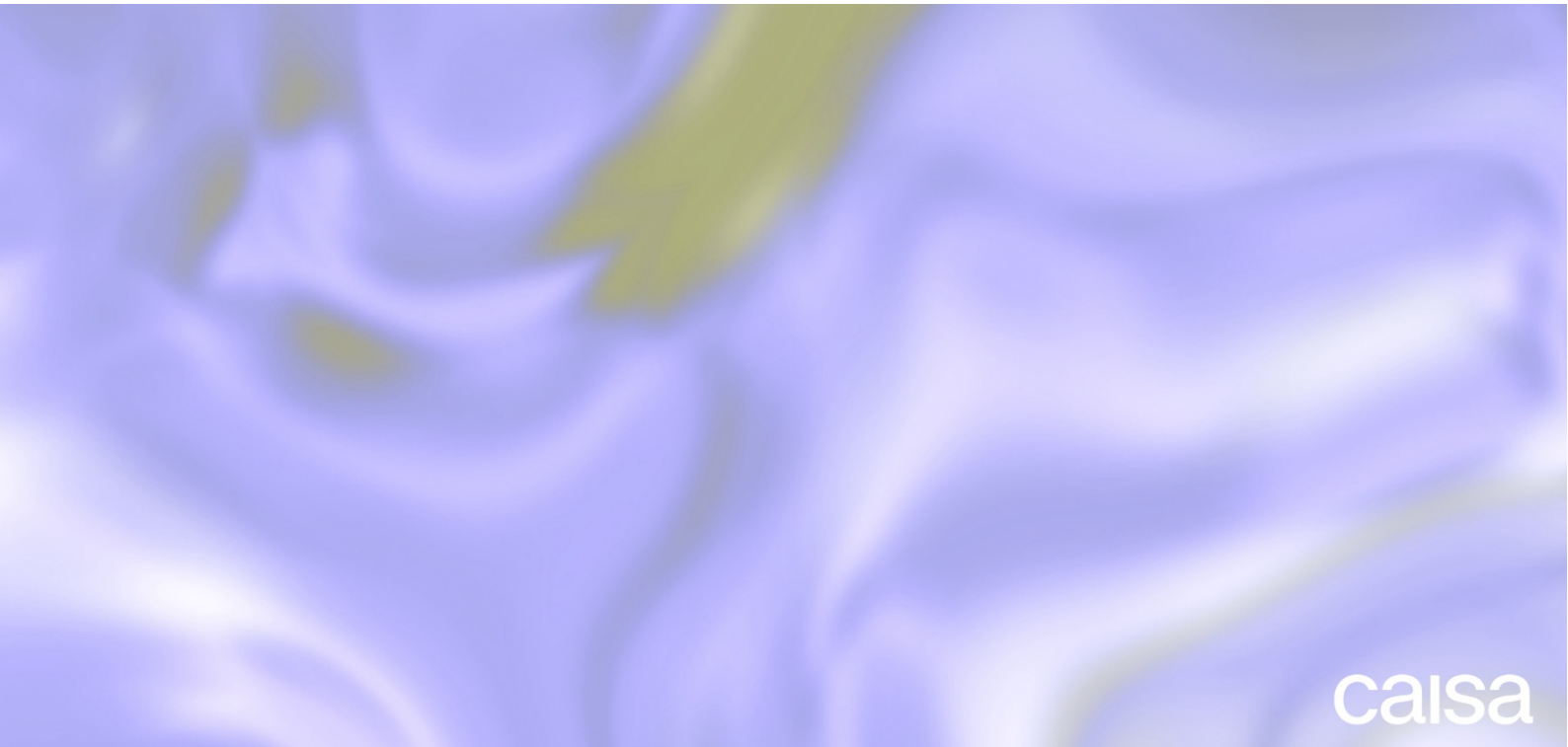
CAISA briefs are part of CAISA's effort to ensure that knowledge and novel insights from within the research community come to empower decision-makers in public and private sectors, and, in turn, society-writ-large when faced with the opportunities and risks posed by rapid technological change. CAISA publishes two types of briefs:

Research briefs present research and evidence-based knowledge about AI and society in an accessible way.

Position briefs express the authors' research based and informed assessment of important challenges related to AI and society.

CAISA's briefs are edited by Anders Søgaard, Professor at the Department of Computer Science at the University of Copenhagen and Chief Scientist at CAISA, and Johannes N. Feldt, Research Assistant at CAISA. All briefs are read by – and receive comments from – at least one external independent researcher before publication.

The authors are responsible for the contents of a CAISA brief.



References

- Abaluck, J., Pless, R., Ravi, N., Sautmann, A., & Schwartz, A. (2026). Does LLM Assistance Improve Healthcare Delivery? An Evaluation Using On-site Physicians and Laboratory Tests. NBER Working Paper No. w34660. https://doi.org/10.3386/w34660
- Agarwal, N., Moehring, A., Rajpurkar, P., & Salz, T. (2025). Combining Human Expertise with Artificial Intelligence: Experimental Evidence from Radiology. NBER Working Paper No. w31422. https://doi.org/10.3386/w31422
- Blease, C.R., Locher, C., Gaab, J., Hägglund, M., & Mandl, K.D. (2024). Generative Artificial Intelligence in Primary Care: An Online Survey of UK General Practitioners. *BMJ Health & Care Informatics* 31(1), e101102. https://pmc.ncbi.nlm.nih.gov/articles/PMC11429366
- Holm, A., Siersma, V., Bjerrum, L., & Cordoba, G. (2020). Availability of point-of-care culture and microscopy in general practice-does it lead to more appropriate use of antibiotics in patients with suspected urinary tract infection? *European Journal of General Practice* 26(1), 175-181. https://doi.org/10.1080/13814788.2020.1853697
- Humlum, A., & Vestergaard, E. (2025). The Unequal Adoption of ChatGPT Exacerbates Existing Inequalities among Workers. *Proceedings of the National Academy of Sciences* 122(1), e2414972121. https://doi.org/10.1073/pnas.2414972121
- Huang, S., M.A. Ribers, & Ullrich, H. (2022). Assessing the Value of Data for Prediction Policies: The Case of Antibiotic Prescribing. *Economics Letters* 213(110360). https://doi.org/10.1016/j.econlet.2022.110360
- Huang, S., & Ullrich, H. (2026). Provider Effects in Antibiotic Prescribing: Evidence from Physician Exits. *Journal of Human Resources* 61(4), 1159-1191. https://doi.org/10.3368/jhr.0523-12900R1
- Huang, S., Schmacker, R., Ullrich, H. (2026). AI Adoption by Human Experts: Evidence from Primary Care Physicians. DIW Discussion Paper Nr. 2168.
- Jørgensen, N.L., Merrild, C.H., Jensen, M.B., Moeslund, T.B., Kidholm, K., & Thomsen, J.L. (2025). The perceptions of potential prerequisites for artificial intelligence in Danish general practice: vignette-based interview study among general practitioners. *JMIR Medical Informatics* 13(1), e63895. https://doi.org/10.2196/63895
- Jørgensen, N.L., Bruun, N.H., Hoffmann Merrild, C., Moeslund, T.B., Kidholm, K., & Thomsen, J.L. (2026). Identifying factors influencing acceptance of artificial intelligence among general practitioners in Danish general practice: a cross-sectional web-based survey study. *Scandinavian Journal of Primary Health Care* 44(1), 2667990. https://doi.org/10.1080/02813432.2026.2667990
- Kanjilal, S., Oberst, M., Boominathan, S., Zhou, H., Hooper, D. C., & Sontag, D. (2020). A decision algorithm to promote outpatient antimicrobial stewardship for uncomplicated urinary tract infection. *Science Translational Medicine* 12(568), eaay5067. https://doi.org/10.1126/scitranslmed.aay5067
- Katonai, G., Arvai, N., & Mesko, B. (2025). AI and primary care: Scoping review. *Journal of Medical Internet Research* 27, e65950. https://doi.org/10.2196/65950
- Mikkelsen, J.G., Sørensen, N.L., Merrild, C.H., Jensen, M.B., & Thomsen, J.L. (2023). Patient perspectives on data sharing regarding implementing and using artificial intelligence in general practice–a qualitative study. *BMC Health Services Research* 23(1), 335. https://doi.org/10.1186/s12913-023-09324-8
- Ribers, M.A., & Ullrich, H. (2023). Machine predictions and human decisions with variation in payoffs and skill: the case of antibiotic prescribing. Berlin School of Economics Discussion Paper No. 27. https://doi.org/10.48462/opus4-5111
- Ribers, M.A., & Ullrich, H. (2024). Complementarities between algorithmic and human decision-making: The case of antibiotic prescribing. *Quantitative Marketing and Economics* 22, 445-483. https://doi.org/10.1007/s11129-024-09284-1
- Saust, L. T., Siersma, V., Lykkegaard, J., Bjerrum, L., & Hansen, M.P. (2022). Diagnosis and antibiotic treatment of urinary tract infections in Danish general practice: a quality assessment. *Antibiotics* 11(12), 1759. https://doi.org/10.3390/antibiotics11121759
- Show Us the Evidence for the Value of Medical AI [Editorial] (2026). *Nature Medicine* 32, 1163. https://doi.org/10.1038/s41591-026-04389-4

CAISA's website

CAISA's LinkedIn