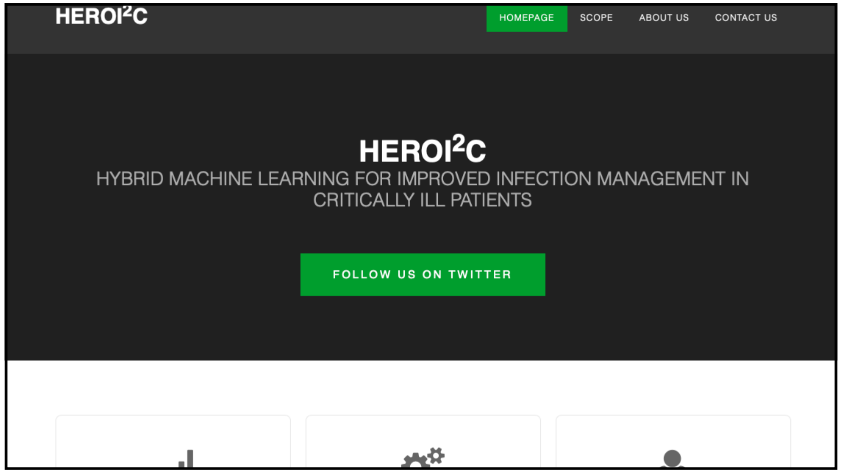
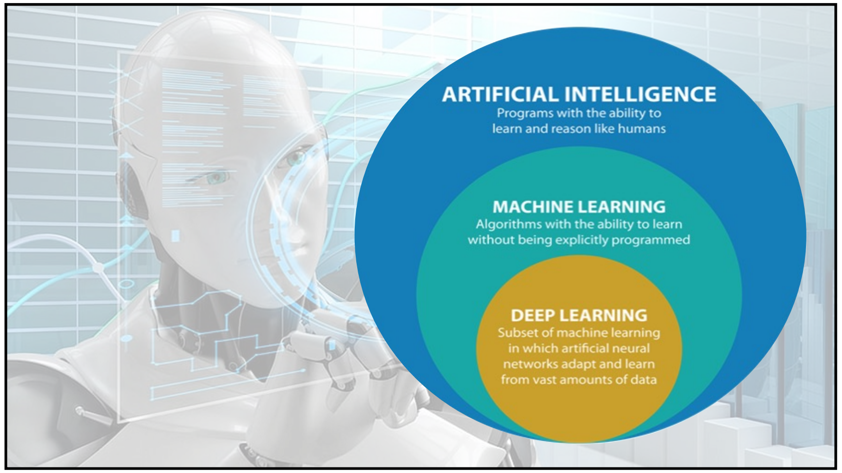

GHENT
UNIVERSITY
THE FUTURE OF
ARTIFICIAL INTELLIGENCE
– SHINY AND BRIGHT?
BY JAN J DE WAELE
CONTACT JAN.DEWAELE@UGENT.BE
TWITTER @CRITICCAREDOC

1


HEROI²C
HOMEPAGE
SCOPE
ABOUT US
CONTACT US
HEROI²C
HYBRID MACHINE LEARNING FOR IMPROVED INFECTION MANAGEMENT IN CRITICALLY ILL PATIENTS
FOLLOW US ON TWITTER

2



3

Applications of AI in critical care

- Disease identification
- Disease evolution prediction
- Disease phenotyping
- Guiding clinical decisions

GHENT UNIVERSITY

4

SYSTEMATIC REVIEW

Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit

Davy van de Sande[1], Michel E. van Genderen[1*], Joost Huiskens[2], Diederik Gommers[1] and Jasper van Bommel[1]

Abstract

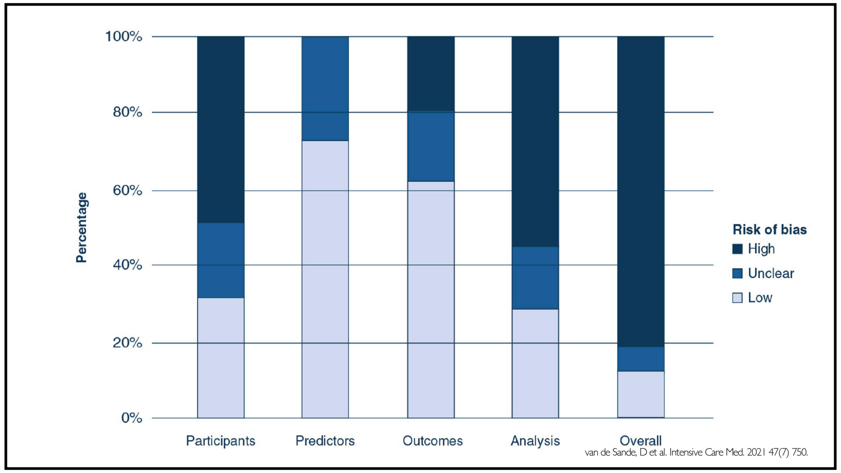
Purpose: Due to the increasing demand for intensive care unit (ICU) treatment, and to improve quality and efficiency of care, there is a need for adequate and efficient clinical decision-making. The advancement of artificial intelligence (AI) technologies has resulted in the development of prediction models, which might aid clinical decision-making. This systematic review seeks to give a contemporary overview of the current maturity of AI in the ICU, the research methods behind these studies, and the risk of bias in these studies.

5

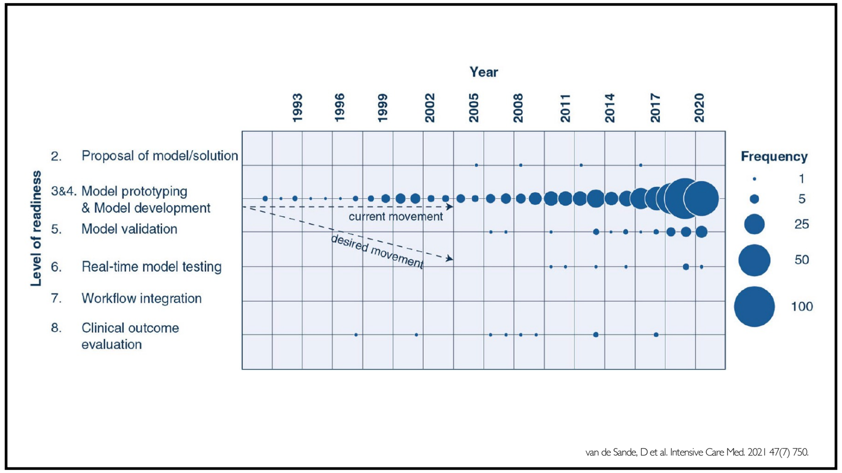
Table 1 Number and proportion (%) of studies according to the study aim and study design

Aim of study	Study design						
	Number (%) of studies with this aim¥	Retrospective*			Prospective observational	Non-randomized clinical trial	Randomized clinical trial
		Internal	External	Non			
Predicting complications	110 (22.2%)	86 (78.2%)	12 (10.9%)	4 (3.6%)	5 (4.5%)	2 (1.8%)	1 (0.9%)
Predicting mortality	102 (20.6%)	92 (90.2%)	9 (8.8%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)
Improving prognostic models/risk scoring system	91 (18.4%)	80 (87.9%)	7 (7.7%)	3 (3.3%)	1 (1.1%)	0 (0%)	0 (0%)
Classifying sub-populations	58 (11.7%)	53 (91.4%)	1 (1.7%)	4 (6.9%)	0 (0%)	0 (0%)	0 (0%)
Determining physiological thresholds	24 (4.9%)	21 (87.5%)	1 (4.2%)	2 (8.3%)	0 (0%)	0 (0%)	0 (0%)
Predicting length of stay	22 (4.4%)	22 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Alarm reduction	21 (4.3%)	20 (95.2%)	1 (4.8%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Predicting medication administration	19 (3.8%)	14 (73.7%)	1 (5.3%)	1 (5.3%)	0 (0%)	2 (10.5%)	1 (5.3%)
Improving mechanical ventilation	16 (3.2%)	13 (81.3%)	0 (0%)	0 (0%)	1 (6.3%)	1 (6.3%)	1 (6.3%)
Assessing clinical notes	13 (2.6%)	9 (69.2%)	1 (7.7%)	1 (7.7%)	0 (0%)	0 (0%)	2 (15.4%)
Predicting readmissions	12 (2.4%)	11 (91.7%)	1 (8.3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Predicting relevance of clinical information	8 (1.6%)	5 (62.5%)	1 (12.5%)	2 (25%)	0 (0%)	0 (0%)	0 (0%)
Assessing videos and images	7 (1.4%)	6 (85.7%)	0 (0%)	0 (0%)	1 (14.3%)	0 (0%)	0 (0%)
Detecting spurious recorded values	6 (1.2%)	6 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Predicting health improvement	5 (1%)	5 (100%)	0 (0%)	0 (0%)	0 (0%)	0(0%)	0 (0%)
Predicting unnecessary lab tests	3 (0.6%)	3 (100%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Total (accounting for duplicates)	494	421 (85.2%)	35 (7.1%)	20 (4%)	8 (1.6%)	5 (1%)	5 (1%)

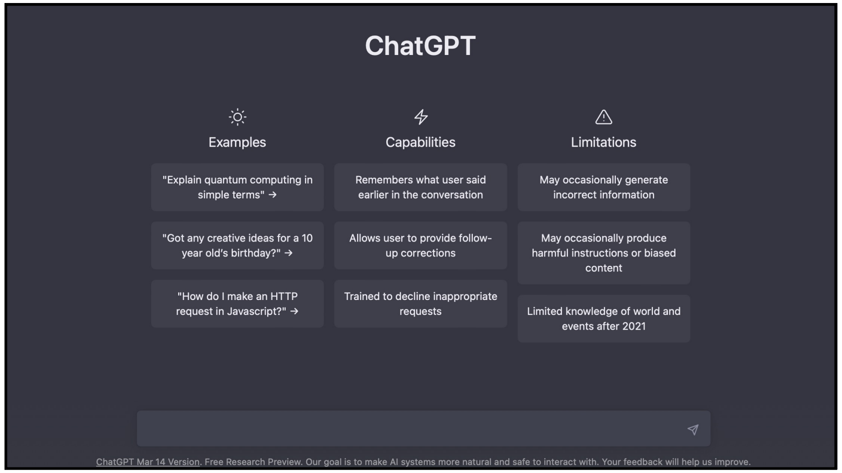
6


100%
80%
60%
40%
20%
0%
Percentage
Participants
Predictors
Outcomes
Analysis
Overall
Risk of bias
High
Unclear
Low
van de Sande, D et al. Intensive Care Med. 2021 47(7) 750.

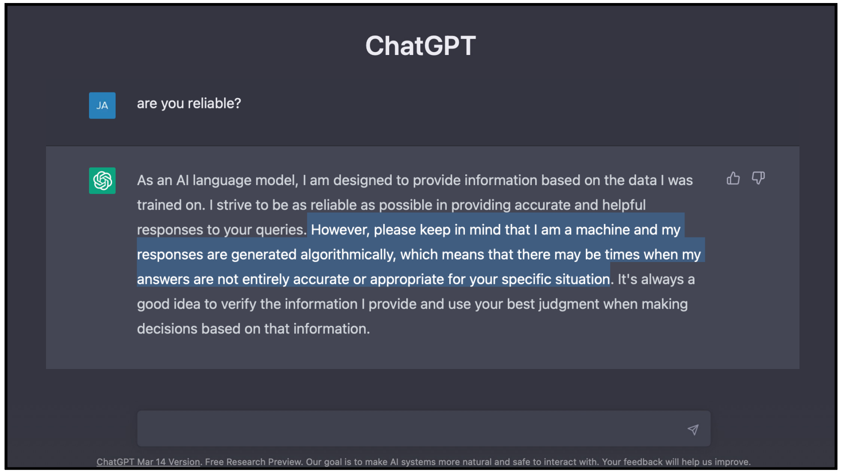
7


Year
1993
1996
1999
2002
2005
2008
2011
2014
2017
2020
Level of readiness
2. Proposal of model/solution
3&4. Model prototyping & Model development
5. Model validation
6. Real-time model testing
7. Workflow integration
8. Clinical outcome evaluation
current movement
desired movement
Frequency
1
5
25
50
100
van de Sande, D et al. Intensive Care Med. 2021 47(7) 750.

8


ChatGPT
Examples
Capabilities
Limitations
"Explain quantum computing in simple terms" →
"Got any creative ideas for a 10 year old's birthday?" →
"How do I make an HTTP request in Javascript?" →
Remembers what user said earlier in the conversation
Allows user to provide follow-up corrections
Trained to decline inappropriate requests
May occasionally generate incorrect information
May occasionally produce harmful instructions or biased content
Limited knowledge of world and events after 2021
ChatGPT Mar 14 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.

9


ChatGPT
JA
are you reliable?
As an AI language model, I am designed to provide information based on the data I was trained on. I strive to be as reliable as possible in providing accurate and helpful responses to your queries. However, please keep in mind that I am a machine and my responses are generated algorithmically, which means that there may be times when my answers are not entirely accurate or appropriate for your specific situation. It's always a good idea to verify the information I provide and use your best judgment when making decisions based on that information.
ChatGPT Mar 14 Version. Free Research Preview. Our goal is to make AI systems more natural and safe to interact with. Your feedback will help us improve.

10

Correspondence

oa

Why we cannot trust artificial intelligence in medicine

George Orwell said, "if thought corrupts language, language can also corrupt thought."[1] Orwell's worry about the totalitarian regimes of his day offers broader insights about how language can obscure thinking and values, including about contemporary medical applications of artificial intelligence (AI). The potential of AI to revolutionise medicine appears vast.[2] Nevertheless, concerns over the unknown and unknowable so-called black boxes of AI have spurred a movement toward building trust in AI.

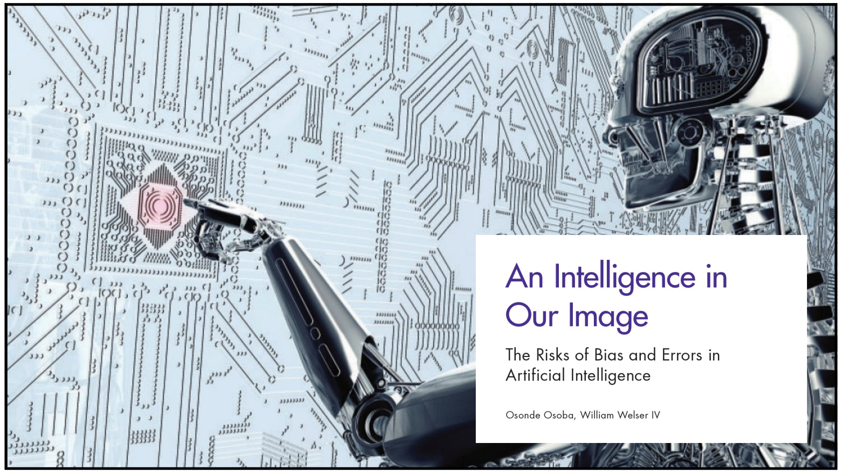
Although well intentioned, applying trust to AI is a category error, mistakenly assuming that AI belongs to a category of things that can be trusted. Trust implies

Embracing trust in AI as if AI were a moral agent also unwittingly fosters diffusion of responsibility.[3] Absolving physicians of blame in times of error while muting praise for wise decisions takes medicine in the wrong direction. Although AI, like a faulty surgical instrument, might be causally implicated, we cannot rightly assign moral responsibility to it. Whether future versions of AI can be regarded as moral agents is only a matter of speculation.[4]

Although in common parlance we certainly speak of trusting machines—eg, trusting cars to get us places—we ought not confuse these colloquialisms with the true meaning of trust. Preserving precision in the usage of trust strikes at the heart of the identity of medicine. As AI increasingly becomes a part of medicine, its proper role should be in supporting effective, empathic, and ethically attentive care for humans,

DeCamp, M et al. Lancet Digit Health. 2019 1(8) e390.

11



12

AI malfunction: data shift

Mismatch between development dataset and clinical application

- ✓ Technology
- ✓ Population/environment eg season, new treatment
- ✓ Behaviour eg patient, change in clinical practice

Physicians need to be aware

Retraining/redesign/recalibrate may be necessary

GHENT UNIVERSITY

13

Research

JAMA Internal Medicine | Original Investigation

External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients

Andrew Wong, MD; Erkin Otles, MEng; John P. Donnelly, PhD; Andrew Krumm, PhD; Jeffrey McCullough, PhD; Olivia DeTroyer-Cooley, BSE; Justin Pestrue, MEcon; Marie Phillips, BA; Judy Konye, MSN, RN; Carleen Penoza, MHSA, RN; Muhammad Ghous, MBBS; Karandeep Singh, MD, MMSc

IMPORTANCE The Epic Sepsis Model (ESM), a proprietary sepsis prediction model, is implemented at hundreds of US hospitals. The ESM's ability to identify patients with sepsis has not been adequately evaluated despite widespread use.

OBJECTIVE To externally validate the ESM in the prediction of sepsis and evaluate its potential clinical value compared with usual care.

DESIGN, SETTING, AND PARTICIPANTS This retrospective cohort study was conducted among 27 697 patients aged 18 years or older admitted to Michigan Medicine, the academic health system of the University of Michigan, Ann Arbor, with 38 455 hospitalizations between

Editorial page 1040

Multimedia

Supplemental content

CME Quiz at jamacmelookup.com and CME Questions page 1148

14

External validation is essential

Table 2. ESM Performance

Model performance	Hospitalization	Time horizons 24 h	12 h	8 h	4 h
Outcome incidence, %	6.6	0.43	0.29	0.22	0.14
Area under the receiver operating characteristic curve (95% CI)	0.63 (0.62-0.64)	0.72 (0.72-0.72)	0.73 (0.73-0.74)	0.74 (0.74-0.75)	0.76 (0.75-0.76)
Positive predictive value (ESM score ≥6), %	12	2.4	1.7	1.4	0.92
No. needed to evaluate (ESM score ≥6)[a]	8	42	59	73	109

GHENT UNIVERSITY

Wong, A et al. JAMA Intern Med. 2021 181(8) 1065.

15

Perspectives

Digital medicine

Artificial intelligence, bias, and patients' perspectives

For more on **Digital medicine** see **Comment** *Lancet* 2016; **388:** 740 and **Perspectives** *Lancet* 2021; **397:** 1254

Some of the most exciting applications of machine learning to medicine involve the kinds of data that cannot be analysed with traditional statistical models: medical imaging, waveforms, and videos. Researchers are training algorithms to take in these complex signals, and output a doctor's interpretation—eg, given a particular retinal fundus photograph, would an ophthalmologist identify diabetic retinopathy? Algorithms based on datasets that pair images or waveforms with "labels" assigned by a doctor have the potential to drive improvements in efficiency and diagnostic accuracy. However, the strength of this approach can also be its weakness: by matching the performance of doctors, algorithms will also incorporate their inherent limitations.

machine learning approach, however, will falter for such a task. By training an algorithm to predict what a radiologist would say about the image—eg, its Kellgren and Lawrence grade—we are also constraining it. We are preventing the algorithm from seeing past the doctor's limitations and biases. The performance of artificial intelligence (AI) algorithms has typically been compared with doctors' performance, but what about patients' experiences?

Research by one of us (ZO), with colleagues, has produced an algorithm trained to predict the knee pain reported by the patient, rather than the x-ray interpretation of the doctor. This approach explained more of all patients' pain compared with standard measures of radiographic severity, and its explanatory power for pain was particularly useful

Obermeyer, Z et al. Lancet. 2021 397(10289) 2038.

16

Explainability and interpretability

White vs black box

Table 1 A non-exhaustive selection of 'black box' definitions

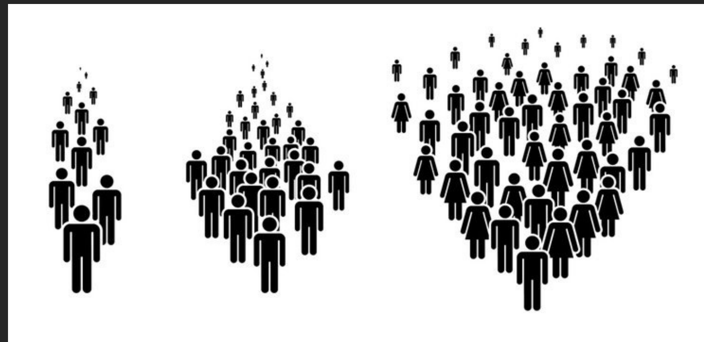
Opaque to experts and programmers	Opaque to non-experts	Opaque to some unspecified person/people
A system's creators have designed and programmed a system, but they cannot explain how it makes decisions.[19]	Only exists when an algorithm is incomprehensible to non-specialists who only see input and output without understanding the process.[20]	A challenge of understanding and explaining how machine learning systems make predictions.[21]
An object whose inner functioning and set-up cannot be known.[22]	Users are not meant to achieve the expertise of a computer scientist, meaning they will not achieve the same level of understanding.[17]	Is simply the mystery of how a given system reaches its various outputs.[23]
Are epistemically opaque, meaning no human or group of humans can survey its inner states.[12]		Systems where the inputs and outputs are known but the internal representations are not understood.[24]

GHENT UNIVERSITY

Wadden, JJ. J Med Ethics. 2021.

17

Algorithm quality



GHENT UNIVERSITY

18

Technological obstacles

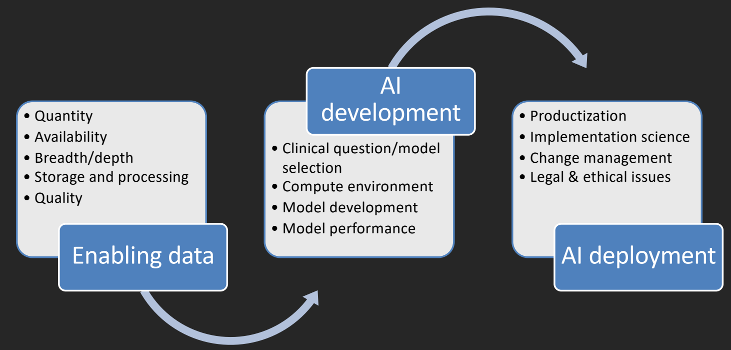
- Data security
- Difficulties in real-time application

- Dashboarding critical
- Monitoring algorithm behaviour

GHENT UNIVERSITY

19

The path to implementing AI



GHENT UNIVERSITY

Mamdani, M et al. Intensive Care Med. 2020

20

Real-world scenario vs experimental setting

Comparison of humans versus mobile phone-powered artificial intelligence for the diagnosis and management of pigmented skin cancer in secondary care: a multicentre, prospective, diagnostic, clinical trial

Scott W Menzies, Christoph Sinz, Michelle Menzies, Serigne N Lo, William Yolland, Johann Lingohr, Majid Razmara, Philipp Tschandl, Pascale Guitera, Richard A Scolyer, Florentina Boltz, Liliane Borik-Heil, Hsien Herbert Chan, David Chromy, David J Coker, Helena Collgros, Maryam Eghtedari, Marina Corral Forteza, Emily Forward, Bruna Gallo, Stephanie Geisler, Matthew Gibson, Amelie Hampel, Genevieve Ho, Laura Junez, Philipp Kienzl, Arthur Martin, Fergal J Moloney, Amanda Regio Pereira, Julia Maria Ressler, Susanne Richter, Katharina Silic, Thomas Silly, Michael Skoll, Julia Tittes, Philipp Weber, Wolfgang Weninger, Doris Weiss, Ping Woo-Sampson, Catherine Zilberg, Harald Kittler

Summary

Background Diagnosis of skin cancer requires medical expertise, which is scarce. Mobile phone-powered artificial

Lancet Digit Health 2023;

GHENT UNIVERSITY

https://doi.org/10.1016/S2589-7500(23)00130-9

21

Real-world scenario vs experimental setting

Comparison of humans versus mobile phone-powered artificial intelligence for the diagnosis and management of pigmented skin cancer in secondary care: a multicentre, prospective, diagnostic, clinical trial

Scott W Menzies, Christoph Sinz, Michelle Menzies, Serigne N Lo, William Yolland, Johann Lingohr, Majid Razmara, Philipp Tschandl, Pascale Guitera, Richard A Scolyer, Florentina Boltz, Liliane Borik-Heil, Hsien Herbert Chan, David Chromy, David J Coker, Helena Collgros, Maryam Eghtedari, Marina Corral Forteza, Emily Forward, Bruna Gallo, Stephanie Geisler, Matthew Gibson, Amelie Hampel, Genevieve Ho,

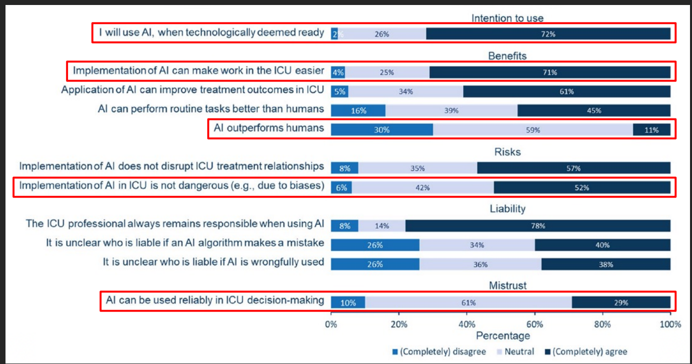
Interpretation The mobile phone-powered AI technology is simple, practical, and accurate for the diagnosis of suspicious pigmented skin cancer in patients presenting to a specialist setting, although its usage for management decisions requires more careful execution. An AI algorithm that was superior in experimental studies was significantly inferior to specialists in a real-world scenario, suggesting that caution is needed when extrapolating results of experimental studies to clinical practice.

GHENT UNIVERSITY

https://doi.org/10.1016/S2589-7500(23)00130-9

22

ICU HCW attitudes towards AI



GHENT UNIVERSITY

van de Sande, D et al. Intensive Care Med. 2022

23

Conclusions

- AI is hot, also in the ICU
- Research is booming, focus on prediction and early diagnosis, phenotyping
- Path to individualized medicine
- Many challenges
- Lack of robustness, risk of bias, data shift, equity, ethical aspects
- Data ownership and regulatory issues
- Critical stance essential

GHENT UNIVERSITY

24

Jan.DeWaele@UGent.be @CriticCareDoc

25