HEIDELBERG
FACULTY OF
MEDICINE

The patient in the loop: autonomy, privacy and responsibility in digital healthcare

PD Dr. Nadia Primc

Institute for the History and Ethics of Medicine
Primc@uni-heidelberg.de

I. Digital Transformation and Disruption of Healthcare

II. AI-based clinical decision and diagnostic support systems and four challenges

Co-funded by
the European Union

Digitalisation of Healthcare (eHealth)

Digitalization: The transition from analog to digital information processing (eHealth)

Covers many different applications:

- Robotics: Surgical robots, social robots
- mHealth applications (apps, websites)
- Monitoring systems (AAL, telemedicine, telenursing)
- Big Data & AI
- Documentation and billing systems, etc.

Adobe.stock

Affects almost every part of the healthcare system, professions, areas of care, and all stakeholders in the healthcare system.

- Digital literacy as a prerequisite for successful digital transformation: not only the safe use of digital technologies, but also their critical reflection; embedded in (continuing) education, training, etc. (cf. new ÄAppo / NKLM 2.0)

The Digital Transformation of the Healthcare System

Digital Transformation (DT) as a widely used „buzzword“

- **Not every instance of digitalization** automatically constitutes a **digital transformation**

  “While all transformation is change, not all change is transformation.” (Gong/Ribiere 2021)

Digital transformation refers to **fundamental change** brought about by digitalization

- Comprehensive (affecting all areas) and far-reaching changes that lead to a **better "outcome"**

Improvement for whom?

→ Benefits for patients and healthcare professionals

19.04.26, 11:26 The outdated machine hampering the fight against Covid-19

BBC

Register Sign In

Home News Sport Business Technology Health Culture Arts Travel Earth Audio Video Live

The outdated machine hampering the fight against Covid-19

6 September 2021

Share Save Add as preferred on Google

Ashifa Kassam

Features correspondent

The use of fax machines in health services have led to delays in Covid-19 tests results (Credit: Alamy)

Major backlogs in processing patient data during the pandemic are forcing the healthcare sector to reassess its relationship with fax machines.

Digitalization and Digital Transformation

Digitalization of Healthcare:

- Process of the gradual implementation of digital technologies with changes and (intended) improvements to healthcare
- To be understood as a **sociotechnical process:** new technologies change and affect/improve healthcare processes (and vice versa)

Digital Transformation:

- Radical change and improvement in the quality of healthcare

"Transformation needs to meet the criteria of the three "Bs" — it must be **Big**, must be **Bold** (i.e., the intensity and degree of change involved), and lead to **Better outcomes**." (Gong/Ribiere 2021

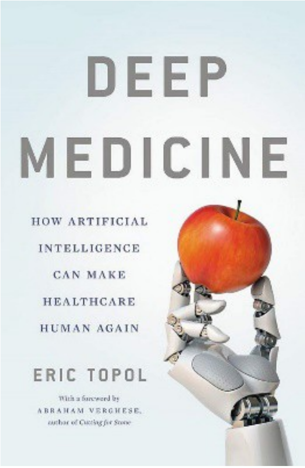
→ Consensus that fundamental changes in healthcare are coming; the open question is what form these will take

hype vs. hope technologie, dystopian vs. utopian visions

Vision of a Digital Transformation of the Healthcare System

Utopia (hope): Realization of 4p Medizin

- Prediction: Forecasting diseases and health risks through improved data analysis
- Prevention: Preventing or intervening early on the basis of data
- Personalization: Collection and analysis of individual health data
- Participation: Involving patients in their treatment, e.g. self-monitoring / self-management (cf. self-knowledge)



Improvement of healthcare through:

- Better networking of stakeholders (professions, sectors)
- Faster and higher-quality data analysis
- Greater involvement of patients in care processes (e.g. empowerment, democratization, adherence)

AI and Digitalization as Disruptive Technologies

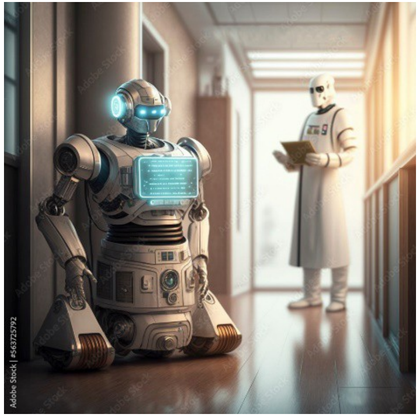
Disruption / the disruptive potential of AI and digital technologies

- Similarly overused term
- frequently left undefined

Often circular reference: disruption ↔ digital transformation

  - disruptive nature of AI and digital technologies as potential to bring about DT *vs.*
  - DT as result of disruptive technologies

More often associated with a negative, dystopian view of digitalization and AI in healthcare



Adobe stock

Disruption and disruptive Technologies

„Disruptive Technologies“, 1995, Clayton M. Christensen (Harvard Business School)

- An economic approach to explaining why leading companies fail to adapt to new innovations
- Theory of failure to adapt to new innovations
- Possible strategy/strategies: early integration of potentially disruptive innovations into one's own business strategy

→ A term shaped by discipline-specific **epistemological and practical goals of economic theory**

Photography by Platon

Disruptive Technologies in Healthcare

Limits of transferability of the concept of disruptive technology, given that the **healthcare system** should not be treated as a **purely economic enterprise**

Lessons to be learned from the concept of disruption for the ethical-legal discussion of AI and digitalization in healthcare

- **potentially disruptive technologies**: developed by interest groups outside of healthcare, each with their own discipline-specific values and objectives (cf. technology push / economization)
- disruption can be counteracted by **incorporating the normative values** inherent to healthcare into the development of technology (participatory approaches, "ethics by design", etc.)

Summary part 1

Digital transformation and **disruption** are **not neutral technical terms**, but as normative concepts, i. e. they **carry value judgements** about what is **good** or **harmful for healthcare**

- **Digital Transformation:** Strengthening of the inherent values and objectives of healthcare
- **Disruption:** Endangerment of the inherent values and objectives of healthcare

I. Digital Transformation and Disruption of Healthcare

II. AI-based clinical decision and diagnostic support systems and four challenges

AI-based systems to support diagnostics and clinical decision-making

Google Deep Mind for breast cancer (McKInney et al. 2020, Nature)

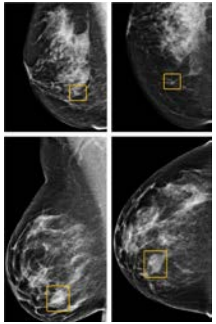
- Typically: 80% of carcinomas are detected, 20% go undetected
- DeepMind: Reduction of false positive results by up to 5.7%, false negative results by up to 9.4% in single-reader assessment (USA)
- Reduction of workload by up to 88% for second radiologists (UK)

IBM Watson for Oncology (WFO)

- Comparison of decisions made by interdisciplinary tumor boards and WFO recommendations
- Differences depending on the study as well as the type & stage of cancer (et al., 2021, Scientific Reports 11)
- Highest agreement for breast cancer Stage I–III (88.99%)

Further examples from dermatology, neuroradiology, transplant medicine, etc.

Google DeepMind detects previously undetected breast carcinomas. Reuters



Opportunities of AI-based systems for supporting diagnostics and clinical decision-making

METIS

- Improvement of disease prediction through more precise diagnostics
- Expansion of therapeutic options available by broader incorporation of scientific evidence
- Realization of the ideal of EBM: evidence-based, data-driven treatment
- Greater personalization through the collection, processing, and analysis of individual health data
- Reduction of the workload of physicians

In Brief

iChoose Kidney for Treatment Options: Updated Models for Shared Decision Aid

Jennifer C. Gander, PhD,[1] Mohua Basu, MPH,[2] Laura McPherson, MPH,[2] Michael D. Garber, MPH,[3] Stephen O. Pastan, MD,[4] Amita Manatunga, PhD,[5] Kimberly Jacob Arriola, PhD,[6] and Rachel E. Patzer, PhD[2,3]

Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records

Hans-Christian Thorsen-Meyer, Annelaura B Nielsen, Anna P Nielsen, Benjamin Skov Kaas-Hansen, Palle Toft, Jens Schierbeck, Thomas Strøm, Piotr J Chmura, Marc Heimann, Lars Dybdahl, Lasse Spangsege, Patrick Hulsen, Kirstine Belling, Søren Brunak, Anders Perner

oa

RESEARCH ARTICLE

Machine Learning Models to Predict Withdrawal of Life-Sustaining Therapy in Patients With Severe Traumatic Brain Injury

Michael Cobler-Lichter,[1] Jessica M. Delamater,[1] Fernanda J.P. Teixeira,[2] Ana M. Reyes,[1] Talia R. Arcieri,[1] Brian Manolovitz,[2] John A. McKeown,[2] Tulay Koru-Sengul,[3] Jonathan Jagid,[4] Joacir Graciolli Cordeiro,[4] Nina Massad,[2] Mohan Kottapally,[2] Amedeo Merenda,[2] Kristine O'Phelan,[2] Nicholas Namias,[1] and Ayham Alkhachroum[2]

Correspondence
Dr. Alkhachroum
axa2610@med.miami.edu

Neurology® 2025;105:e214249. doi:10.1212/WNL.0000000000214249

Challenges of AI-CDSS/DSS for patient-centered care

- Black-box problem
- Automation bias
- Responsibility gap
- Privacy and data
- [....]

The Black-Box Problem and different forms of opacity

= difficulty of understanding how an AI system arrives at its outputs.

You can observe what goes in (input) and what comes out (output), but the internal reasoning process is **opaque**!

→ cannot be interrogated or explained in terms that humans can understand and evaluate

→ accuracy vs. transparency/explainability (XAI)

Different forms of opacity

- **technical opacity**: the model is mathematically complex (algorithmic, representational)
- **epistemic opacity**: knowledge about the model is limited or absent (training data, epistemic, temporal)
- **social/institutional opacity**: access or communication is restricted (intentional, user-facing, e.g. proprietary systems)

Technical vs. Epistemic Opacity — A Closer Look

Technical opacity: intrinsic property of the model itself

- e.g. deep learning model with high accuracy in detection of breast cancer from mammograms, that outperforms radiologists in clinical trials, but nobody can explain how
- Its decision emerges from hundreds of millions of weights interacting across many layers

Epistemic opacity: property of our knowledge about the model, not of the model itself.

- It arises from gaps, absences, and uncertainties in what has been recorded, documented, or studied [e.g. training data opacity: We don't know exactly what went into training, documented health records might have absorbed biases, gaps, etc.]
- It is in principle reducible: better records, and more research could close the gap.

Would full access and perfect documentation solve the problem?

✓ Yes: it's epistemic opacity.

✗ No: it's technical opacity.

Black Box Problem and why it matters ....

e.g. 4 Principles of biomedical ethics

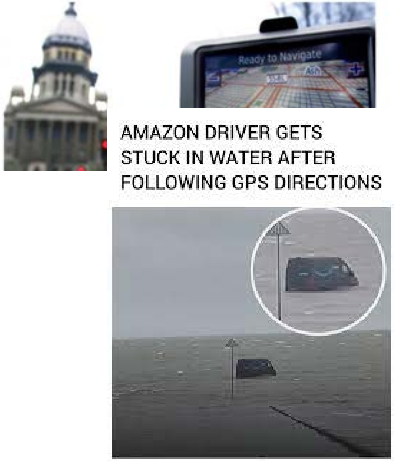
- limits the **capacity for informed consent** of patients/surrogates (autonomy)
- limits ability of doctors **to fulfil their duties of beneficence and non-maleficence** (state of professional incompetence, e.g. responsibility)
- limits ability **to assess fairness of decisions** and to detect discriminations (justice)

Automation bias

- tendency of humans to **over-rely on automated systems**, deferring to their outputs even when those outputs are wrong, and even when the person has **sufficient knowledge** or **information to recognise** the error.
- Automation Bias and AI: systems amplify automation bias in ways older automated tools did not, for several reasons.
  - AI outputs often come with **confident, fluent explanations** (e.g. LLMs), which make them feel more authoritative and reasoned than a simple alert or number would.
  - **XAI** tools, paradoxically, can **worsen automation bias**: when an AI highlights a region of a scan or lists the features driving a prediction, the explanation can feel like verification
  - high-stakes and time-pressured environments (where AI decision support is most desirable) are also the ones where cognitive offloading is most tempting and most dangerous.


m+ Tourists blindly follow their GPS and end up driving on a SKI SLOPE in Andorra

By OLIVIA ALLHUSEN, FOREIGN NEWS REPORTER and GAÏA SVENSTEDT
PUBLISHED: 10:31 BST, 13 January 2026 | UPDATED: 17:17 BST, 13 January 2026



When something goes wrong: the responsibility gap

Automation bias strikes and something goes wrong: who is responsible?

- Clinician, who should have known it better?
- Developer(s)?
- Hospital that puchased the system?

Many hands problem: distribution (and diffusion) of responsibility amongst many actors involved

- Developers developed the system, but cannot anticipate every situation/case that the system would encounter
- Deployer (hospital) chose the system, but typically doesn‘t understand its internal working, relied on it in good faith (often following regulatory guidance)
- User (clinician) used the system and was present when the harm occured, but didn‘t design trhe system, relied on it in good faith (often following regulatory guidance)
- AI system itself cannot be a moral or legal subject. It has no intentions, no interests, no capacity for guilt or reform.

Responsibility gap as a structural hindrance to implementation of AI-based systems?

The gap exists in medicine without AI too!

Medicine already operates with distributed and sometimes indeterminate responsibility, e.g. multidisciplinary teams, handovers, complex causal chains between decision and harm.

→ Clinicians and institutions have developed practical norms for navigating this without waiting for philosophical resolution.

→ AI introduces a new version of an old problem rather than an entirely novel one.

Legal and insurance frameworks will need to adapt (e.g. AI Act)

Privacy, Data and Clinical AI — The Core Tension

Machine learning systems (esp. deep learning) require large, diverse, and richly labelled datasets to perform reliably

- structural pressure towards the **aggregation of patient data at scale**, across institutions, across time, and increasingly across national borders

Health data one of the most sensitive categories of private information

- classical informed consent: patient is told, at a specific moment, what will be done with their data, by whom, for what purpose, and with what risks

  → fits poorly with the realities of AI development/secondary use of clinical data

- Risk of re-identification rises with amount of data and development of technologies (anonymization/pseudonymization)

nature communications

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature communications › articles › article

Article | Open access | Published: 23 July 2019

Estimating the success of re-identifications in incomplete datasets using generative models

Luc Rocher, Julien M. Hendrickx & Yves-Alexandre de Montjoye

Nature Communications 10, Article number: 3069 (2019) | Cite this article

From privacy protection to data solidarity?

- critique of the individualistic model that dominates current data governance: most data governance frameworks are designed to protect the individuals from whom data originates.
- reorientation from individual to collective governance. Solidarity-based data governance, by strengthening collective control and ownership of data

→ to ensure that the benefits and costs of digital practices are borne collectively and fairly (cf. Public Health Ethics)


THE LANCET
Digital Health

This journal | Journals | Publish | Clinical | Global health | Multimedia | Events | About

COMMENT · Volume 4, Issue 11, E773-E774, November 2022 · Open Access | Download Full Issue

Data solidarity: a blueprint for governing health futures

Barbara Prainsack [a] · Seliem El-Sayed [b] · Nikolaus Forgó [c] · Łukasz Szoszkiewicz [d] · Philipp Baumer [b]

Affiliations & Notes ∨ Article Info ∨

Summary

Digital transformation of healthcare should be understood as the aspiration to promote patient-centered care, i.e. to keep the patient in the loop

A patient who is in the loop in a meaningful sense

- must be able to understand, how the tools being used work and what their limitations are (black-box challenge)
- must be able to trust that the clinical recommendations she receives reflect genuine professional judgment, not mechanical algorithmic deference (automation bias)
- must be able to ask who is responsible, when sth goes wrong and to receive a meaningful answer (responsibility gap)
- must be able to make informed choices about how her data is used, but should also be addressed as a member of a community (privacy vs. data solidarity)

Brief (!) Survey

Next lecture: Mobilising citizen science in support of health research and health services provision

23/04/2026 at 13:00

Dr. Anna Berti SUMAN