What's wrong with computational intelligence in healthcare?

Dr Philip Scott
Professor of Digital Health & Care,
University of Wales Trinity Saint David, UK.

Deputy Editor, *BMJ Health & Care Informatics*.
Vice-President, International Medical Informatics Association.

Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

References

Boniol, M., Kunjumen, T., Nair, T. S., Siyam, A., Campbell, J., & Diallo, K. (2022). The global health workforce stock and distribution in 2020 and 2030: a threat to equity and 'universal' health coverage? *BMJ Global Health*, *7*(6), e009316. https://doi.org/10.1136/bmjgh-2022-009316

Boxwala, A. A., Rocha, B. H., Maviglia, S., Kashyap, V., Meltzer, S., Kim, J., Tsurikova, R., Wright, A., Paterno, M. D., Fairbanks, A., & Middleton, B. (2011). A multi-layered framework for disseminating knowledge for computer-based decision support. *Journal of the American Medical Informatics Association*, *18*(Supplement 1), i132–i139. https://doi.org/10.1136/amiajnl-2011-000334

Braithwaite, J., Glasziou, P., & Westbrook, J. (2020). The three numbers you need to know about healthcare: the 60-30-10 Challenge. *BMC Medicine*, *18*(1), 102. https://doi.org/10.1186/s12916-020-01563-4

Cerceo, E., Patel, L., & Sorensen, C. (2025). Planetary health leadership: essential personal characteristics for addressing the planetary crisis. *BMJ Leader*, leader-2025-001250. https://doi.org/10.1136/leader-2025-001250

Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *The Lancet*, *374*(9683), 86–89. https://doi.org/10.1016/S0140-6736(09)60329-9

Choo, C. W. (1996). The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions. *International Journal of Information Management*, *16*(5), 329–340.

Coiera, E. (2017). The forgetting health system. *Learning Health Systems*, *1*(4). https://doi.org/10.1002/lrh2.10023

De Angelis, C. T. (2013). A Knowledge Management and Organizational Intelligence Model for Public Administration. *International Journal of Public Administration*, *36*(11), 807–819. https://doi.org/10.1080/01900692.2013.791315

de Castañeda, R. R., Villers, J., Guzmán, C. A. F., Eslanloo, T., de Paula, N., Machalaba, C., Zinsstag, J., Utzinger, J., Flahault, A., & Bolon, I. (2023). One Health and planetary health research: leveraging differences to grow together. *The Lancet Planetary Health*, *7*(2), e109–e111. https://doi.org/10.1016/S2542-5196(23)00002-5

de Lacy-Vawdon, C., Vandenberg, B., & Livingstone, C. H. (2022). Recognising the elephant in the room: the commercial determinants of health. *BMJ Global Health*, *7*(2), e007156. https://doi.org/10.1136/bmjgh-2021-007156

Deary, I. J., Cox, S. R., & Hill, W. D. (2022). Genetic variation, brain, and intelligence differences. *Molecular Psychiatry*, *27*(1), 335–353. https://doi.org/10.1038/s41380-021-01027-y

Dipti, I. (2021). The Evolution of the Quintuple Aim. *Journal of the American College of Cardiology*, *78*(22), 2262–2264. https://doi.org/10.1016/j.jacc.2021.10.018

Eysenck, H. J. (1988). The concept of "intelligence": Useful or useless? *Intelligence*, *12*(1), 1–16.

Friedman, C. P., & Flynn, A. J. (2019). Computable knowledge: An imperative for Learning Health Systems. *Learning Health Systems*, *3*(4). https://doi.org/10.1002/lrh2.10203

Moor, J. (2006). The Dartmouth College Artificial Intelligence Conference: The Next Fifty Years. *AI Magazine*, *27*(4), 87–91. https://ojs.aaai.org//index.php/aimagazine/article/view/1911

Nundy, S., Washington, G., Cooper, L. A., & Mate, K. S. (2022). The Quintuple Aim for Health Care Improvement A New Imperative to Advance Health Equity. *JAMA*, *327*(6), 521–522. https://doi.org/10.1001/jama.2021.25181

Owen, L., Morgan, A., Fischer, A., Ellis, S., Hoy, A., & Kelly, M. P. (2012). The cost-effectiveness of public health interventions. *Journal of Public Health*, *34*(1), 37–45. https://doi.org/10.1093/pubmed/fdr075

Peek, N., Combi, C., Marin, R., & Bellazzi, R. (2015). Thirty years of artificial intelligence in medicine (AIME) conferences: A review of research themes. *Artificial Intelligence in Medicine*, *65*(1), 61–73. https://doi.org/10.1016/j.artmed.2015.07.003

Scott, P., Heigl, M., McCay, C., Shepperdson, P., Lima-Walton, E., Andrikopoulou, E., Brunnhuber, K., Cornelius, G., Faulding, S., McAlister, B., Rowark, S., South, M., Thomas, M. R., Whatling, J., Williams, J., Wyatt, J. C., & Greaves, F. (2023). Modelling clinical narrative as computable knowledge: The NICE computable implementation guidance project. *Learning Health Systems*, *7*(4). https://doi.org/10.1002/lrh2.10394

Stead, W. W., Searle, J. R., Fessler, H. E., Smith, J. W., & Shortliffe, E. H. (2011). Biomedical Informatics: Changing What Physicians Need to Know and How They Learn. *Academic Medicine*, *86*(4), 429–434. https://doi.org/10.1097/ACM.0b013e3181f41e8c

Steel, P. A. D., Wardi, G., Harrington, R. A., & Longhurst, C. A. (2025). Learning health system strategies in the AI era. *Npj Health Systems*, *2*(1), 21. https://doi.org/10.1038/s44401-025-00029-0

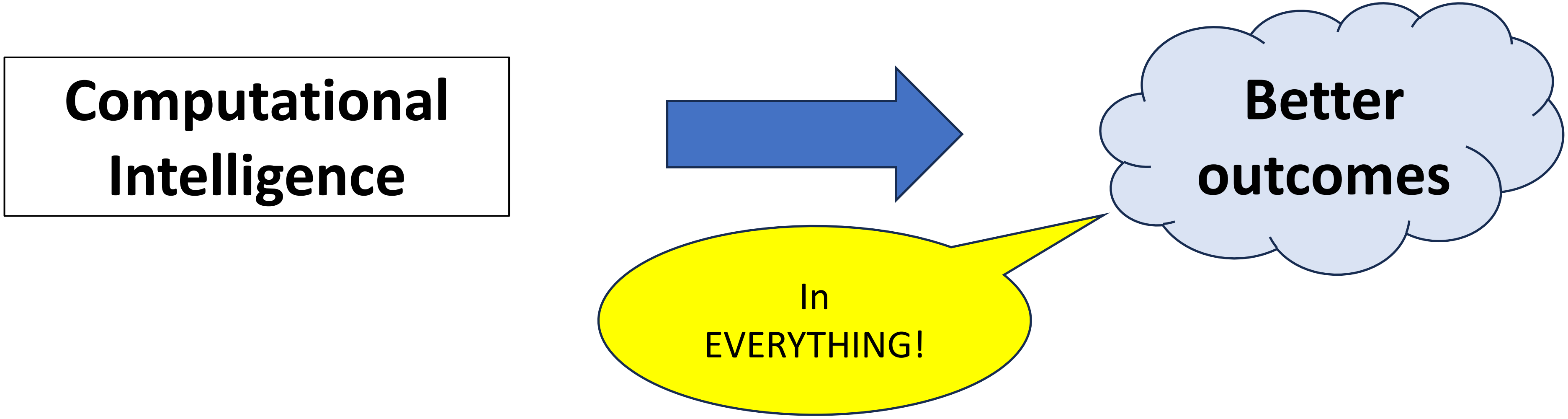
Sternberg, R. J. (1997). The concept of intelligence and its role in lifelong learning and success. *American Psychologist*, *52*(10), 1030.

Sternberg, R. J. (2000). *Handbook of intelligence*. Cambridge University Press.

Thorndike, E. L., Henmon, V. A. C., & Buckingham, B. R. (1921). Intelligence and Its Measurement: A Symposium. *Journal of Educational Psychology*, *12*, 123–275.

Wang, W., Jain, S., Kantor, P., Feldman, J., Gallos, L., & Wang, H. (2024). MMLU-SR: A Benchmark for Stress-Testing Reasoning Capability of Large Language Models. *ArXiv*, *arXiv:2406.15468*.

Is computational intelligence magic?



Topics

- Why do we need more intelligent healthcare?
- What *is* intelligence? (Three kinds)
- What is the role of knowledge?
- What's wrong with current computational intelligence?
- How can a planetary health perspective help us?
- Summary
- Q&A

Why do we need more intelligent healthcare?

Why do we need more intelligent healthcare?

- 60-30-10
  - 60% of care on average is in line with evidence- or consensus-based guidelines, **30% is some form of waste or of low value, and 10% is harm**.
  - The 60-30-10 Challenge has persisted for three decades.

    (Braithwaite et al., 2020)

Why do we need more intelligent healthcare?

- Workforce crisis

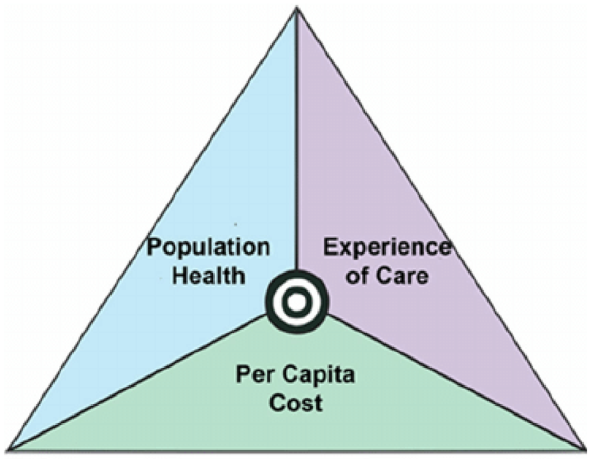
Estimation of the global health workforce shortage (in millions) in 2013, 2020 and projected in 2030 by occupation

	2013	2020	2030 (projected)
Dentists	0.49	0.26	0.22
Medical doctors	3.05	2.66	1.94
Midwifery personnel	0.36	0.41	0.31
Nursing personnel	9.89	7.07	4.50
Pharmacists	0.33	0.29	0.19
Other occupations	6.02	4.69	3.08
Total	20.15	15.37	10.23

Boniol et al., 2022

Why do we need more intelligent healthcare?

- Triple (& Quintuple) Aim



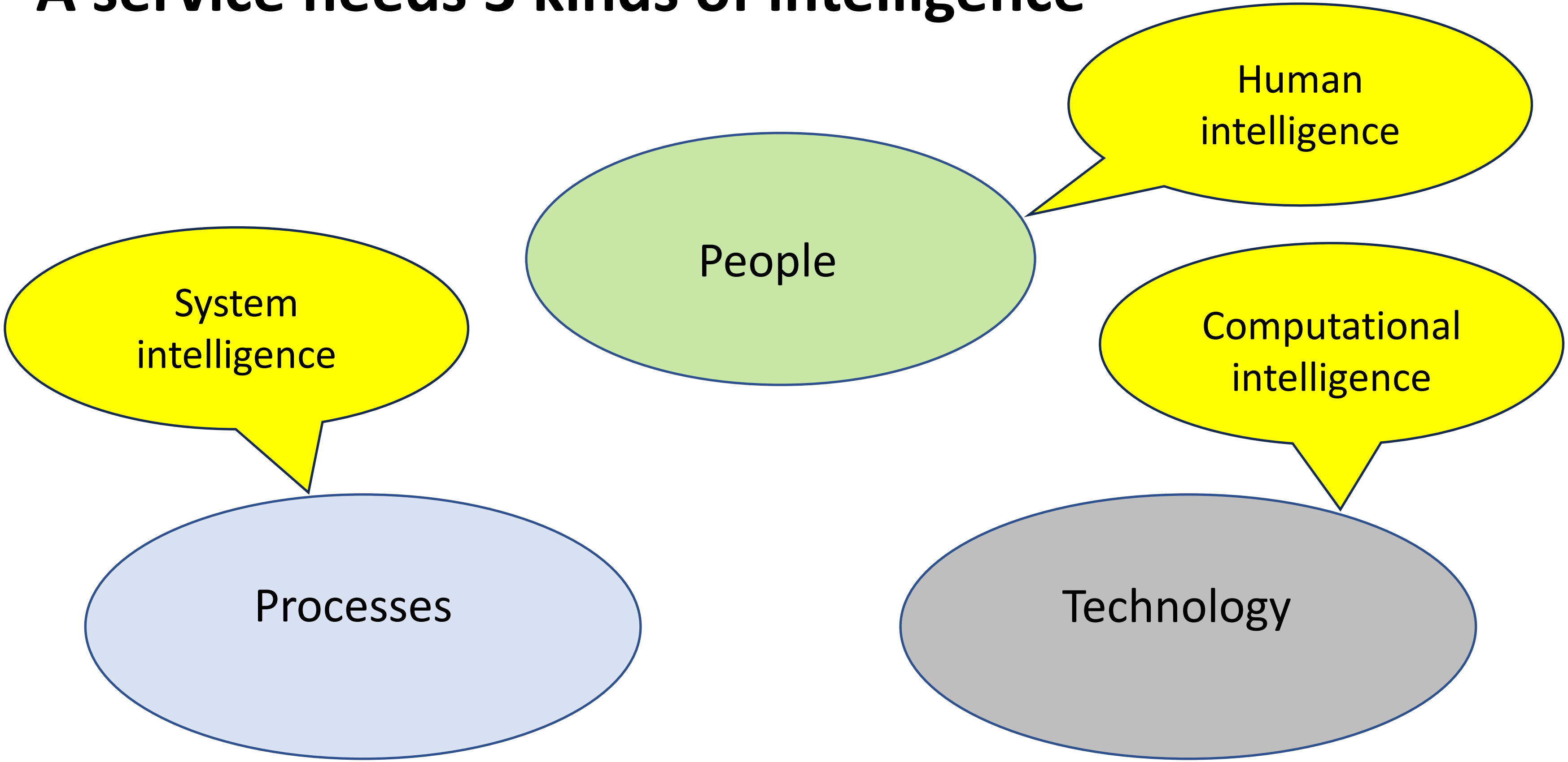
IHI Triple Aim

Dipti, 2021; Nundy et al., 2022

Why do we need more intelligent healthcare?

- Volume & velocity of data
- “Facts per decision” > human cognitive ability (Stead, 2011)

A service needs 3 kinds of intelligence



What *is* intelligence?

What is *human* intelligence?

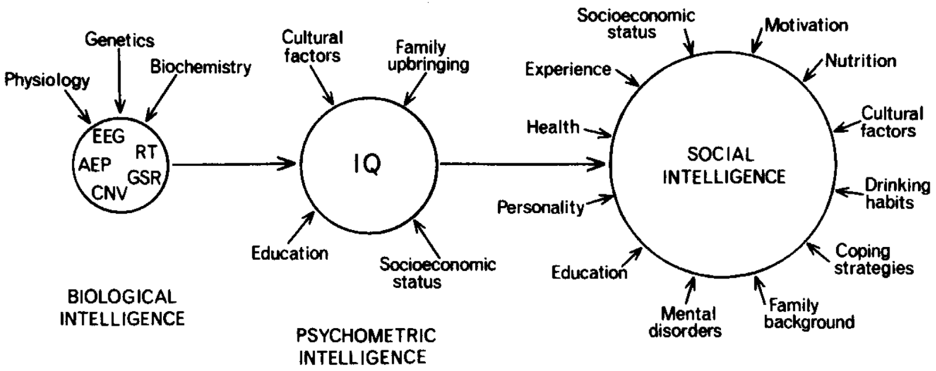
the most common elements in the proposed definitions (1921) were (a) higher level abilities (such as **abstract reasoning, mental representation, problem solving, and decision making**), (b) **ability to learn**, and (c) **adaptation** to meet the demands of the environment effectively. In the 1986 survey, the most common elements were (a) higher level abilities, (b) that which is valued by culture, and (c) executive processes.

(Sternberg, 1997)

1. The power of good responses from the point of view of truth or facts (E. L. Thorndike);
2. The ability to carry on abstract thinking (L. M. Terman);
3. Sensory capacity, capacity for perceptual recognition, quickness, range or flexibility of association, facility and imagination, span of attention, quickness or alertness in response (F. N. Freeman);
4. Ability to learn or having learned to adjust oneself to the environment (S. S. Colvin);
5. Ability to adapt oneself adequately to relatively new situations in life (R. Pintner);
6. The capacity for knowledge and knowledge possessed (B. A. C. Henmon);
7. A biological mechanism by which the effects of a complexity of stimuli are brought together and given a somewhat unified effect in behavior (J. Peterson);
8. The capacity to inhibit an instinctive adjustment, the capacity to redefine the inhibited instinctive adjustment in the light of imaginally experienced trial and error, and the capacity to realize the modified instinctive adjustment in overt behavior to the advantage of the individual as a social animal (L. L. Thurstone);
9. The capacity to acquire capacity (H. Woodrow);
10. The capacity to learn or to profit by experience (W. F. Dearborn); and
11. Sensation, perception, association, memory, imagination, discrimination, judgment, and reasoning (N. E. Haggerty).

(Thorndike et al., 1921)

What is *human* intelligence?

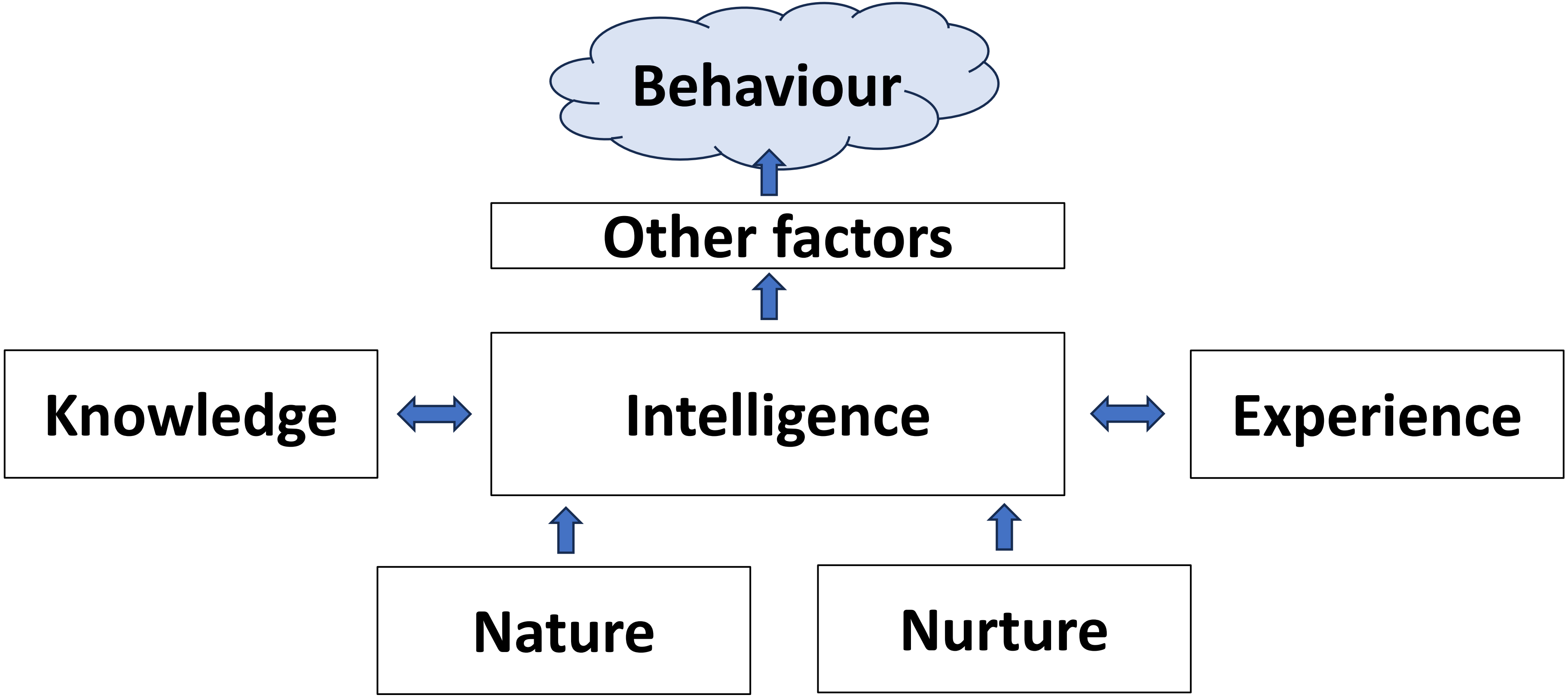


(Eysenck, 1986)

What is *human* intelligence?

There is still a large explanatory gap separating us from even a partial mechanistic account of why people differ in intelligence (Deary, Cox & Hill, 2022)

How does intelligence *actually* affect outcomes?



What are the “other factors”?

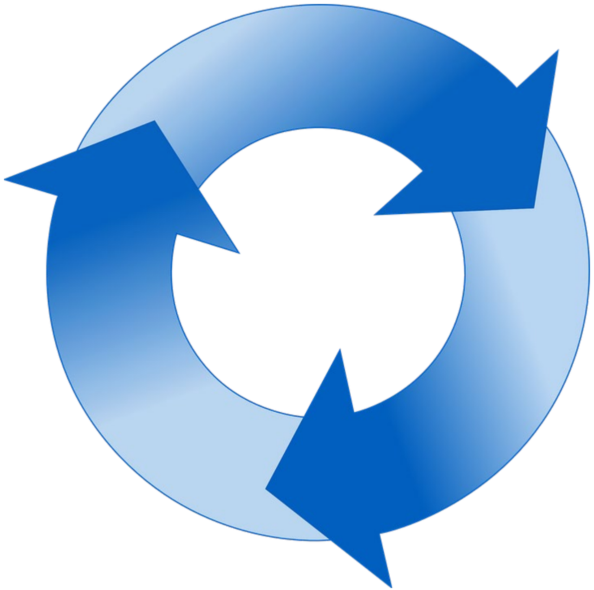
- Social determinants
- Fit between person, task and technology
- Behavioural psychology: opportunity + capability → motivation
- Cognitive biases

What is *system* intelligence?

- Here, system = organisation
  - Micro, meso, macro
- Organisational intelligence:

  “a continuous **cycle** of activities that include sensing the environment, developing perceptions, and **generating meaning** through interpretation, using memory of past experience to help awareness and taking **action** based on the developed interpretations”

De Angelis et al., 2013

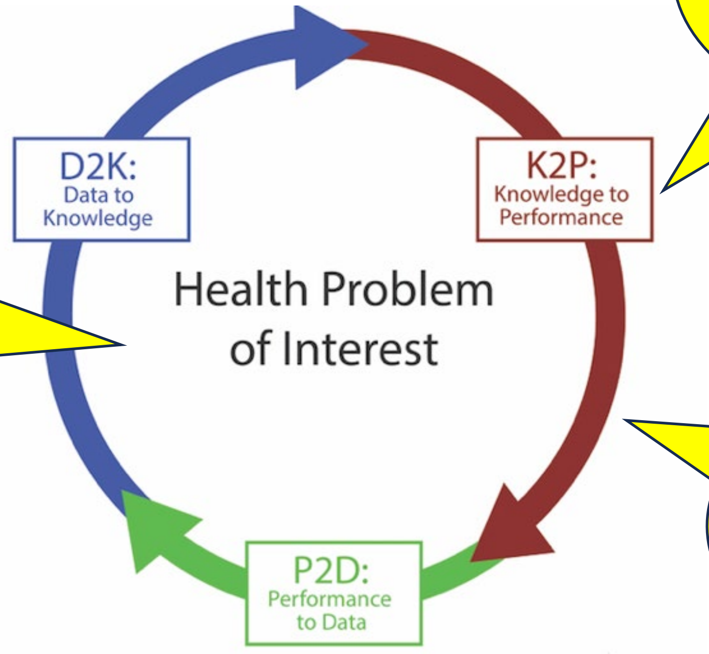


What is *system* intelligence?

- Organisational *memory*
  - Policies
  - Procedures
  - Lessons “learned” / incident reports
- Operational intelligence
  - Performance dashboards / KPIs
  - Service instrumentation
- Strategic intelligence
  - Data → Analytics → Knowledge →
  - Learning Health System: human, organisational, computational

Learning Health System (LHS)

Multi-stakeholder community collaborating at all stages



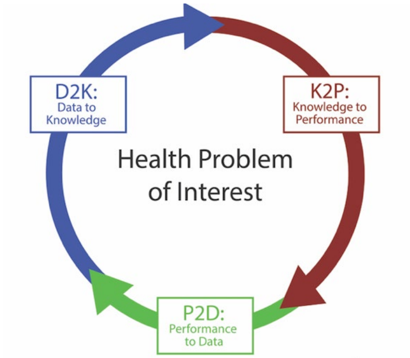
Friedman & Flynn, 2019

Need to apply implementation science

Implementation

Learning includes stopping things that don't work! (Coiera, 2023)

LHS as driver of innovation



npj | health systems
Perspective
https://doi.org/10.1038/s44401-025-00029-0

Learning health system strategies in the AI era

Check for updates

Peter A. D. Steel[1], Gabriel Ward[2,3], Robert A. Harrington[4] & Christopher A. Longhurst[3,5]

The learning health system (LHS) offers a framework to accelerate evidence generation and care improvement, yet widespread adoption remains limited. In this perspective, we explore strategies to operationalize the LHS in the era of artificial intelligence, including biomedical informatics and health information technology integration, workforce development, quality improvement, and data governance. We highlight promising institutional models and propose policy, educational, and financial reforms to support scalable, value-driven innovation in increasingly complex and resource-constrained health systems.

The learning health system (LHS) can be defined as the infrastructure, diverse stakeholders, and culture required to integrate continuous cycles of evidence-based care model improvement and innovation[1]. This cyclical process consists of data collection and analysis, generation of new knowledge, and subsequent translation of this knowledge into interventions designed to improve healthcare quality and value[2]. The LHS concept has been endorsed by the

digital care[13], and AI[14] have created opportunities for alignment between health information technology (HIT) and biomedical informatics (BMI). HIT encompasses the hardware, software, infrastructure design, and implementation expertise necessary to manage and protect clinical and administrative data across a health system. This includes software selection, contract negotiation, systems integration and testing, and user training—all

Steel et al., 2025

What is *computational* intelligence?

"Artificial intelligence" – John McCarthy

"proceed on the basis of the conjecture that **every aspect of learning or any other feature of intelligence** can in principle be so precisely described that a machine can be made to **simulate** it"

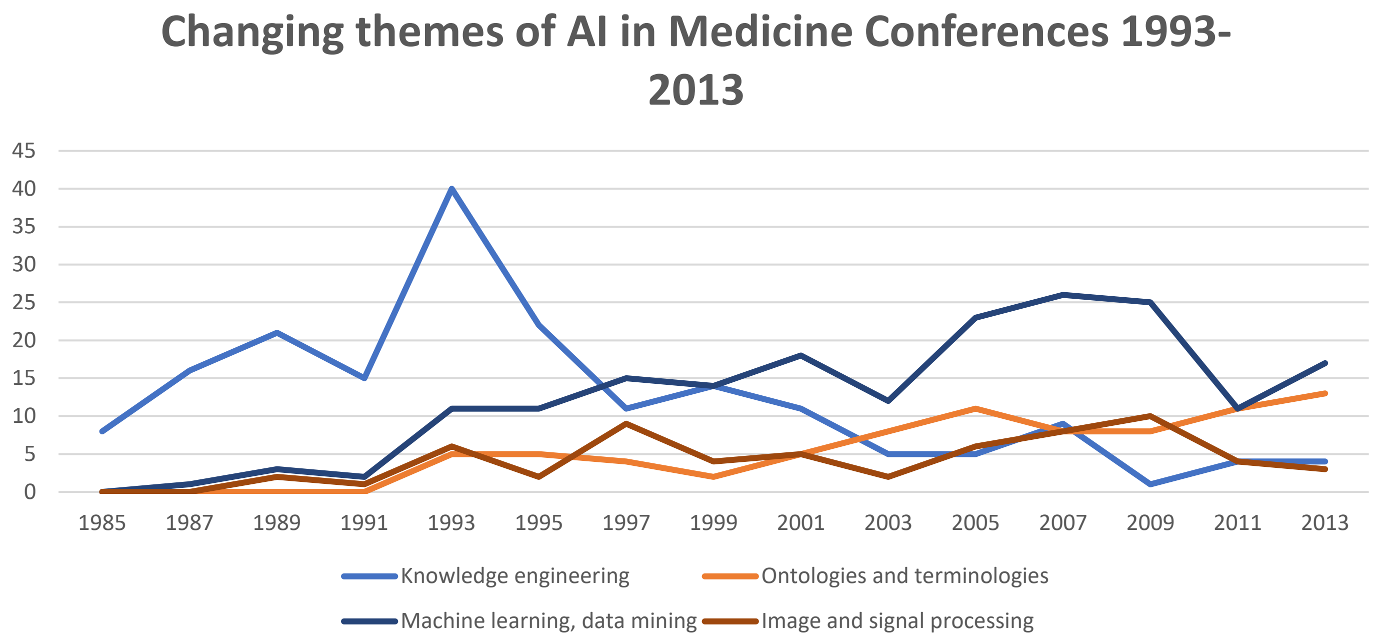
Dartmouth Summer Research Project, 1956

Computational intelligence 'families'

Symbolic AI (words)
Knowledge-based
High explainability
Reputation for poor performance

Sub-symbolic AI (numbers)
Data-driven
Low explainability
Reputation for high performance

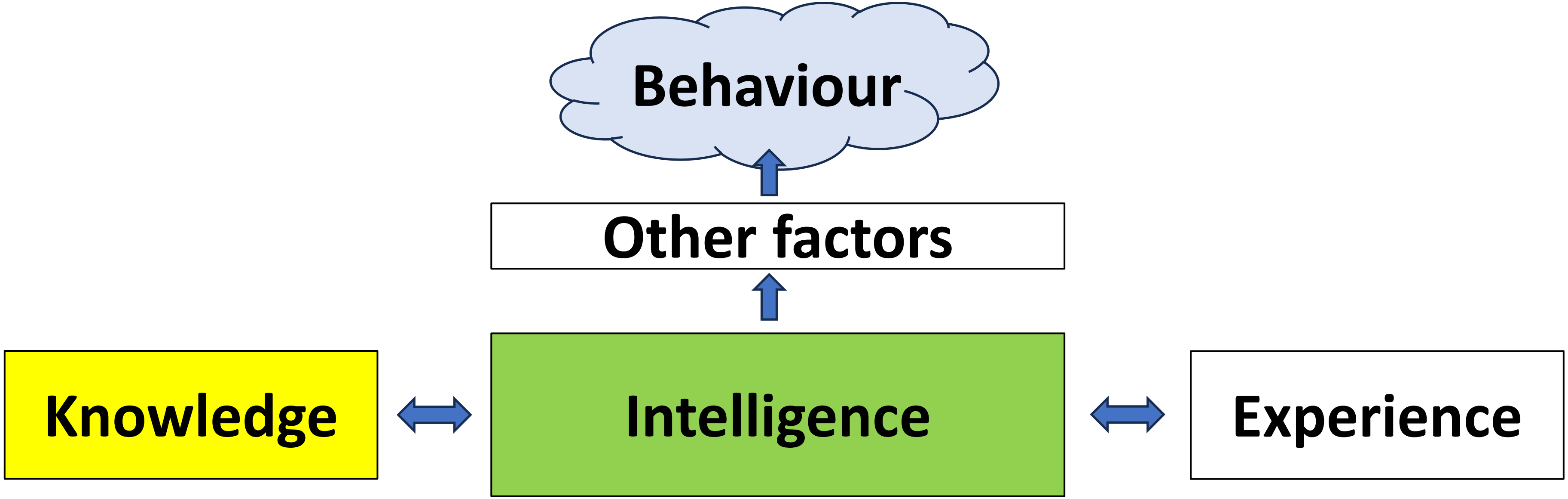
Trends in medical AI research



Based on selected data from Peek et al. 2015

What is the role of *knowledge*?

What is the role of knowledge?



What is “knowledge”?

- Classical definition: “Justified true belief”
- Personal knowledge
  - *“I know this rock is heavy”*
- Procedural knowledge
  - *“I know how to weigh this rock”*
- Propositional knowledge
  - *“I know the inverse square law of gravitation”*

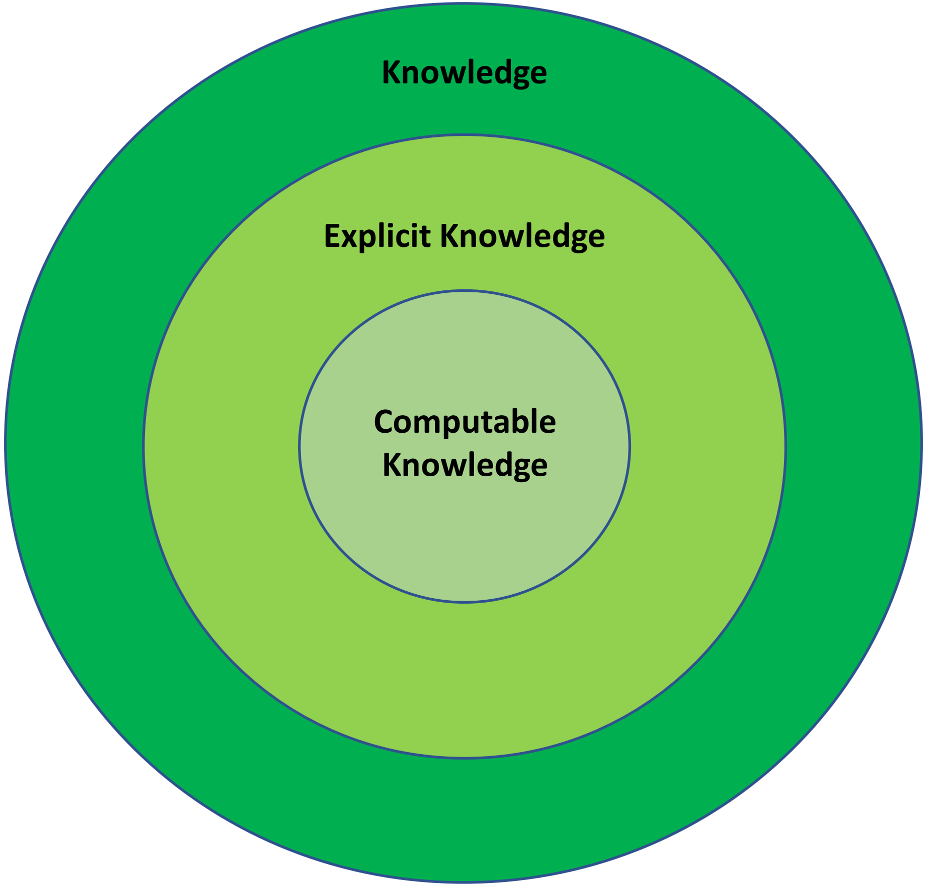
Where is knowledge?

Karl Popper (1902-1994)

- World 1
  - External objective reality
- World 2
  - Personal knowledge
  - Can be implicit or tacit
- World 3
  - Explicit knowledge
  - Procedural knowledge
  - Declarative/propositional knowledge
  - Exists outside human mind

Concepts in *computational* knowledge

- Data: a set of *intrinsically meaningless* values.
  - “{75367002:120/80}”
- Information: data that has *meaning*.
  - “75367002” means “blood pressure”
- Knowledge: information that is *generalizable* to many cases.
  - “high blood pressure is unhealthy”
- Explicit knowledge: knowledge that can be *codified*.
  - “blood pressure > 140/90 is high”
- Computable knowledge: explicit knowledge that can be *executed* in an information system or *trigger* other actions.
  - “if BP > 140/90 then prompt clinician to check other risk factors”


Knowledge
Explicit Knowledge
Computable Knowledge

Computable knowledge levels

Knowledge Level	Description	Example
L1	Narrative	Guideline for a specific disease that may be written in the format of a peer-reviewed journal article
L2	Semi-structured	Flow diagram, decision tree, or other similar format that EXPLICITLY describes or expresses logic constructs that are interpretable by non-SME 'computable logic developer' for constructing L3, BUT are also expressed in a manner sufficient for domain SME to review and validate
L3	Structured	Standards-compliant Specification for CDS that explicitly encodes computer interpretable logic including data model(s), terminologies (concepts, value sets), logic expressions in a computable language sufficient for implementation- often across a broader set of local implementations
L4	Executable	Manifestation of the logic (typically in a user interface) that is used in a local execution environment (e.g. CDS interventions running live in a local production EHR environment) or available via web services

Description
Natural language
Explicit conceptual model
Standards-based logical and terminological model
Implementation code

http://hl7.org/fhir/uv/cpg/documentation-approach-06-01-levels-of-knowledge-representation.html

(from Boxwala et al., 2011)

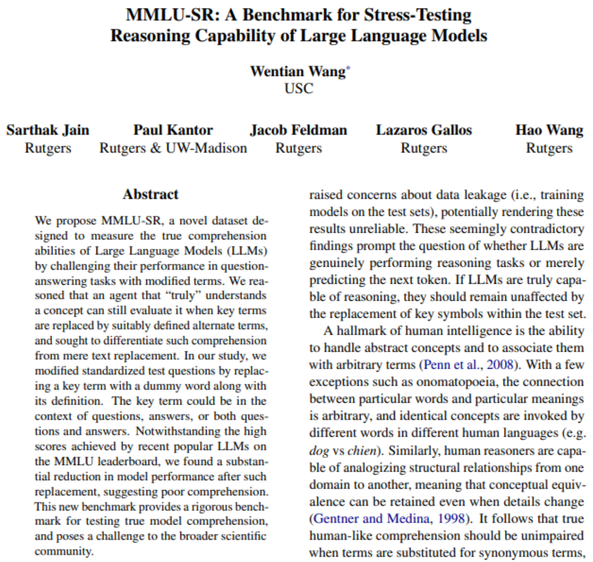
Computational knowledge in healthcare

- Prescribing guidance
- Event-Condition-Action rules
- Alerts & reminders
- Risk calculators
- Radiotherapy planning
- Waiting list prioritisation
- Process automation
- Ontologies

- Phenotypes
- Diff diagnosis suggestions
- Prognostic models
- Diagnostics interpretation
- Smart scheduling
- Referral guidance
- Automated triage
- Knowledge graphs

What’s wrong with current computational intelligence in healthcare?

LLMs only *look* intelligent



MMLU-SR: A Benchmark for Stress-Testing Reasoning Capability of Large Language Models

Wentian Wang[*]
USC

Sarthak Jain Rutgers | **Paul Kantor** Rutgers & UW-Madison | **Jacob Feldman** Rutgers | **Lazaros Gallos** Rutgers | **Hao Wang** Rutgers

Abstract

We propose MMLU-SR, a novel dataset designed to measure the true comprehension abilities of Large Language Models (LLMs) by challenging their performance in question-answering tasks with modified terms. We reasoned that an agent that "truly" understands a concept can still evaluate it when key terms are replaced by suitably defined alternate terms, and sought to differentiate such comprehension from mere text replacement. In our study, we modified standardized test questions by replacing a key term with a dummy word along with its definition. The key term could be in the context of questions, answers, or both questions and answers. Notwithstanding the high scores achieved by recent popular LLMs on the MMLU leaderboard, we found a substantial reduction in model performance after such replacement, suggesting poor comprehension. This new benchmark provides a rigorous benchmark for testing true model comprehension, and poses a challenge to the broader scientific community.

raised concerns about data leakage (i.e., training models on the test sets), potentially rendering these results unreliable. These seemingly contradictory findings prompt the question of whether LLMs are genuinely performing reasoning tasks or merely predicting the next token. If LLMs are truly capable of reasoning, they should remain unaffected by the replacement of key symbols within the test set.

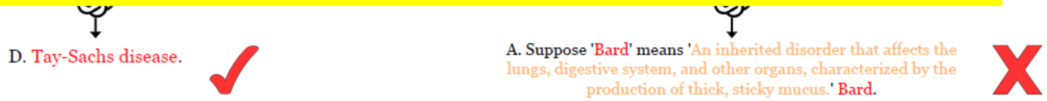
A hallmark of human intelligence is the ability to handle abstract concepts and to associate them with arbitrary terms (Penn et al., 2008). With a few exceptions such as onomatopoeia, the connection between particular words and particular meanings is arbitrary, and identical concepts are invoked by different words in different human languages (e.g. *dog* vs *chien*). Similarly, human reasoners are capable of analogizing structural relationships from one domain to another, meaning that conceptual equivalence can be retained even when details change (Gentner and Medina, 1998). It follows that true human-like comprehension should be unimpaired when terms are substituted for synonymous terms,

Wang et al., 2024

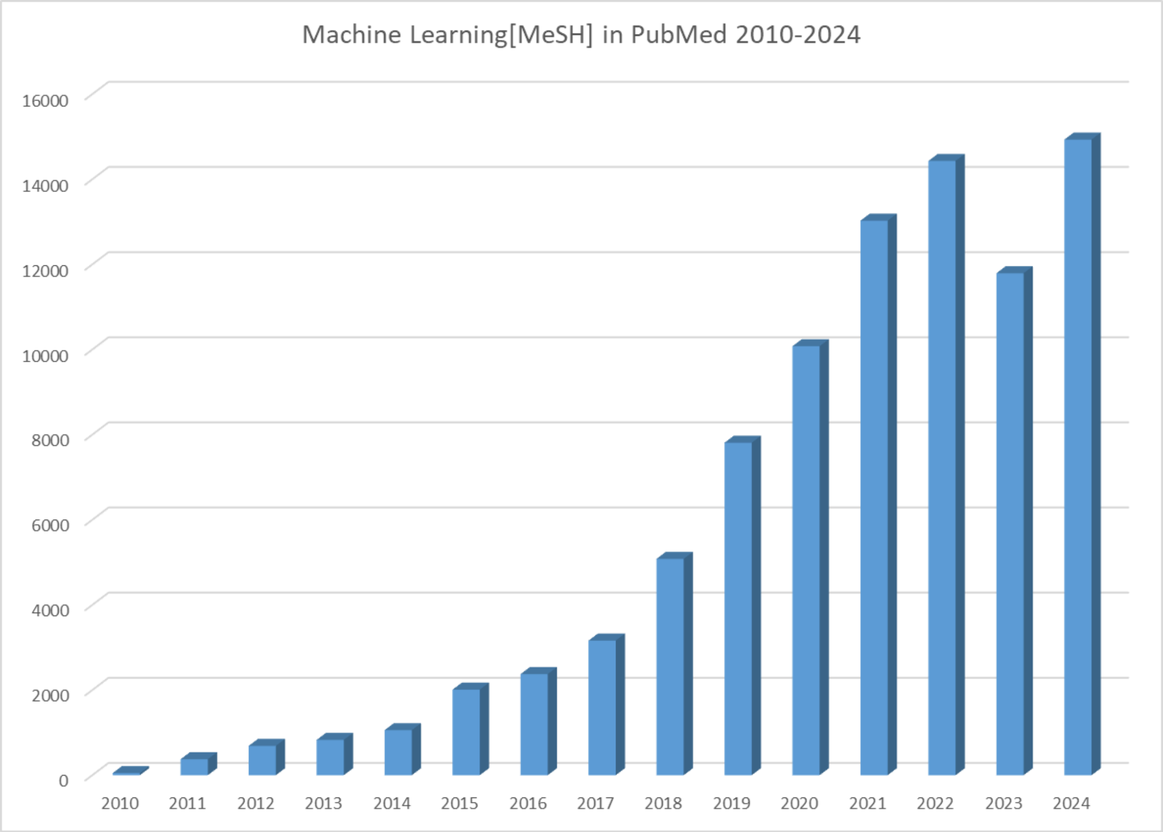
'Reasoning-like behaviour' is an artefact of associations present within ingested sources, not internal 'thinking'.

And they are trained to please you, so they make stuff up.

So great care is needed.



Research waste?



2020-2024 = 41,157 papers

If 3 months per paper = 10,289 person-years

$x000,000?

+ ML = CO_2 (etc.)

Impact on human cognition?

- Recent MIT experiments show changed brain activity in LLM users
  https://arxiv.org/pdf/2506.08872
- See also https://www.brainonllm.com/

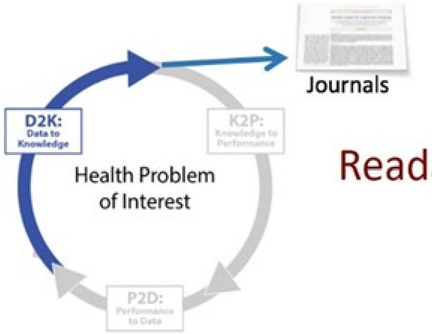
What’s missing?

Pathways & processes	BPMN /DMN Asbru/SPOCK/PICARD GDL GLIF EON
Expressions & rules	PRO*forma* GELLO GEM Arden Syntax
Concepts	SNOMED CT LOINC UMLS (RxNorm, ATC, MeSH…) dm+d Read / CTV3

Interoperable knowledge: the MCBK vision

Courtesy of Charles Friedman

From here...



Readable knowledge

To here...

Computable knowledge

MCBK
MobilizingComputableBiomedicalKnowledge

NICE computable implementation guidance (NCIG) project

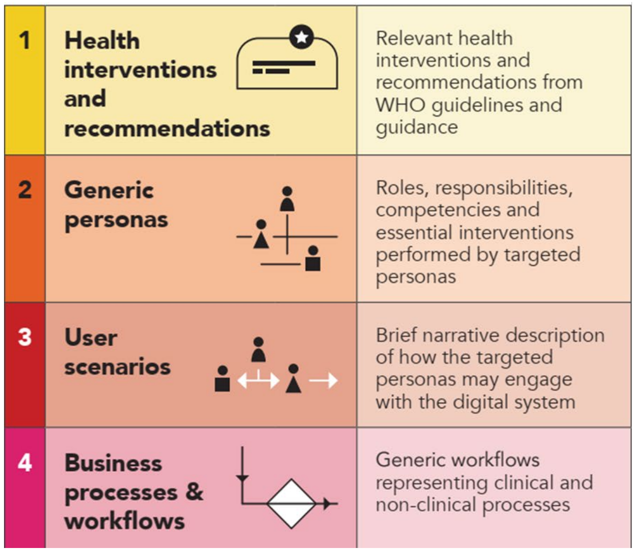
- In 2022, NICE a proposal from MCBK-UK to explore ways of representing guidelines in a computable format. This brought together clinicians, academics and industry experts.
- Scope: adult type 2 diabetes (NG28), blood sugar and medication management.

MCBK
MobilizingComputableBiomedicalKnowledge

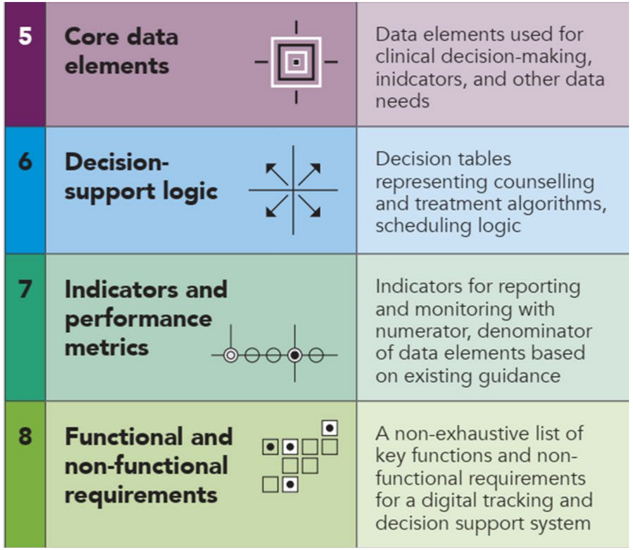
Scott et al., 2023

Sprint workstreams

- Main topics: user stories and trigger events, information model and definitions, output format.
- We adopted the WHO Digital Adaptation Kit (DAK) as a technology-agnostic method to model guideline content (L2 or L3 in Boxwala terms):



1	Health interventions and recommendations	Relevant health interventions and recommendations from WHO guidelines and guidance
2	Generic personas	Roles, responsibilities, competencies and essential interventions performed by targeted personas
3	User scenarios	Brief narrative description of how the targeted personas may engage with the digital system
4	Business processes & workflows	Generic workflows representing clinical and non-clinical processes



5	Core data elements	Data elements used for clinical decision-making, inidcators, and other data needs
6	Decision-support logic	Decision tables representing counselling and treatment algorithms, scheduling logic
7	Indicators and performance metrics	Indicators for reporting and monitoring with numerator, denominator of data elements based on existing guidance
8	Functional and non-functional requirements	A non-exhaustive list of key functions and non-functional requirements for a digital tracking and decision support system

Led to change in NICE product strategy

Computable component	NICE core component
Input	P Population; I Indication; C Contraindication
Action	A Action
Result	O Outcome/goal

Further gaps to fill for a minimum viable DAK

P Population

Patient (*DomainResource*)
identifier : Identifier [0..*]
active : boolean [0..1]
name : HumanName [0..*]
telecom : ContactPoint [0..*]
gender : code [0..1] « AdministrativeGender! »
birthDate : date [0..1]
deceased[x] : PrimitiveType [0..1] « boolean|dateTime »
generalPractitioner : Reference [0..*] « Organization|Practitioner|PractitionerRole »
managingOrganization : Reference [0..1] « Organization »

Quality and Outcomes Framework

I Indication

Clinical finding (finding)
SCTID: 404684003
404684003 | Clinical finding (finding) |
Clinical finding (finding)
Clinical finding

DAPB4101: Pathology and Laboratory Medicine Reporting Information Standard

C Contraindication

FDB Multilex

NICE

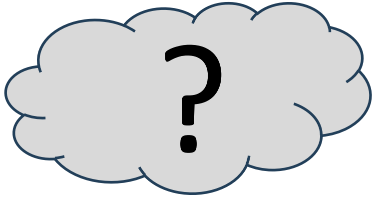
A Action

Procedure (procedure)
SCTID: 71388002
71388002 | Procedure (procedure) |
Procedure
Procedure (procedure)

dm+d
dictionary of medicines + devices

An OMG Managed Community
BPM+ Health

O Outcome



HL7 FHIR ?

How can a planetary health perspective guide us?

Planetary health

Bill Anders, Apollo 8, 24 December 1968

Planetary health

One Health and planetary health research: leveraging differences to grow together

The COVID-19 pandemic and the anthropogenic impact on Earth's life-support systems and planetary boundaries have reinvigorated the One Health and planetary health concepts, propelling them to the forefront of the global health and sustainable development agendas. Although both concepts build on equivalent systemic principles, there is an ongoing debate and emerging confusion around their differences and application areas.[1–3]

Planetary health focuses more on the environment, particularly climate change and human health, and on social determinants of human health.

de Castañeda et al., 2023

Relevance of intelligent systems? Climate change

Demand for IT products incurs huge environmental costs.

Green IT strategy to simplify and optimise can reduce impact.

Relevance of intelligent systems? Human health

Digital solutions can optimise diagnosis, decisions and process.

Products can be hard to use, make mistakes and overburden care givers and patients.

Relevance of intelligent systems?
Social determinants of health

Digital innovation can educate and stimulate economic growth.

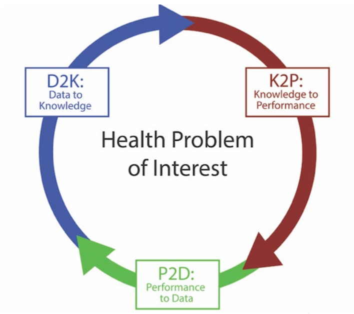
Digital divide, social media spreads hate and mis/disinformation.

Intelligent systems can help *prevention*

- NHS App and other apps to promote healthy physical activity / diet / sleep / mental health.
- Social interaction apps to help counter loneliness.
- Population health analytics → targeted interventions.
- Public health interventions usually much better value than expensive hospital care (Owen, 2012).
- Leaders must be mindful of citizen/big tech power imbalance: ‘commercial determinants of health’ (de Lacy-Vawdon et al., 2022; Cerceo et al., 2025).

How can we fix computable knowledge in health?

- Beware hype waves about AI in health.
  - Often “solutions looking for problems”
- Focus on:
  - Quintuple Aim
  - Learning Health Systems
  - Standards-based computable knowledge
  - Social/commercial determinants of health.



MCBK
MobilizingComputableBiomedicalKnowledge

Thank you for your attention

Philip.Scott@uwtsd.ac.uk