

20 OTTOBRE 2023



AI E DINTORNI

COME TESTARE LA FIDUCIA VERSO UN SERVIZIO DIGITALE

INTELLIGENZA UMANA E ARTIFICIALE

1. **Intelligenza umana e intelligenza artificiale**
2. **The bitter lesson**
3. **Il problema della explainability**
4. **Nel needfinding si possono esplorare trust e explainability?**

L'AI e intelligenza umana si somigliano?

Trends in Cognitive Sciences

CellPress
REVIEWS

Series: Machine Behavior

Opinion

Understanding Human Intelligence through Human Limitations

Thomas L. Griffiths^{1,2,*}

Recent progress in artificial intelligence provides the opportunity to ask the question of what is unique about human intelligence, but with a new comparison class. I argue that we can understand human intelligence, and the ways in which it may differ from artificial intelligence, by considering the characteristics of the kind of computational problems that human minds have to solve. I claim that these problems acquire their structure from three fundamental limitations that apply to human beings: limited time, limited computation, and limited communication. From these limitations we can derive many of the properties we associate with human intelligence, such as rapid learning, the ability to break down problems into parts, and the capacity for cumulative cultural evolution.

Highlights

Humans are limited in time, computation, and communication, defining a set of computational problems that human intelligence has to solve.

Considering the structure of these computational problems can help us to understand why human minds have some of the characteristics that they do.

The solutions to these problems involve mathematical formalisms such as Bayesian inference and meta-learning, rational meta-reasoning, and distributed algorithms, which may be particularly relevant to cognitive science.

1) Ottenere un modello da pochi dati

- Gli umani usano il ragionamento induttivo (e talvolta sbagliano)
- Le AI non sono progettate per imparare da pochi dati (sbagliano più degli umani)
- Le attuali possibilità di computazione ci permettono di avere grandi risorse computazionali



L'AI può avere sempre risorse illimitate?

- Anche le AI devono essere sostenibili
- Le AI sono molto utili per l'esplorazione spaziale, ma non hanno accesso a risorse energetiche illimitate



The "bitter lesson"

- Getting more data is often easier, and more successful, than trying to engineer good inductive biases, something that has been termed the 'bitter lesson' (<http://www.incompleteideas.net/InIdeas/BitterLesson.html>).



2) Dividere un
problema
complesso in
sotto-task



2) Computazioni parallele?

- Gli algoritmi di deep learning e machine learning non devono necessariamente suddividere gli obiettivi in sotto-task
- Questo aspetto però peggiora l'explainability



Il cubo di Rubick

SIAM J. DISCRETE MATH.
Vol. 27, No. 2, pp. 1082–1105

© 2013 Society for Industrial and Applied Mathematics

THE DIAMETER OF THE RUBIK'S CUBE GROUP IS TWENTY*

TOMAS ROKICKI[†], HERBERT KOCIEMBA[‡], MORLEY DAVIDSON[§], AND JOHN
DETHRIDGE[¶]

Abstract. We give an expository account of our computational proof that every position of Rubik's Cube can be solved in 20 moves or less, where a move is defined as any twist of any face. The roughly 4.3×10^{19} positions are partitioned into about two billion cosets of a specially chosen subgroup, and the count of cosets required to be treated is reduced by considering symmetry. The reduced space is searched with a program capable of solving one billion positions per second, using about one billion seconds of CPU time donated by Google. As a byproduct of determining that the diameter is 20, we also find the exact count of cube positions at distance 15.

Key words. group theory, algorithm performance, Rubik's Cube

AMS subject classifications. 20-04, 05C12, 20B40

DOI. 10.1137/120867366

- Gli umani più rapidi usano circa 50 mosse per risolvere il cubo di Rubick
- Per ottenere questo risultato occorre memorizzare un centinaio di mosse (tante, ma ricordabili) da applicare in specifiche situazioni
- Il metodo umano è basato su subtask e ha dunque una specifica struttura
- In questo modo è molto più facilmente rappresentabili e insegnabile



*I bias nel
ragionamento
umano*

E l'AI?



● *Take home
message*

*Intelligenza umana e
artificiale sono
diverse*

Risorse

L'AI HA RISORSE ILLIMITATE
MENTRE L'INTELLIGENZA
UMANA NASCE DALLA
NECESSITÀ DI SERVIRSI DI
RISORSE LIMITATE

Limiti

SI PUÒ INSEGNARE A UNA AI UN
RAGIONAMENTO INDUTTIVO?

*il problema
explainability*

LIMITARE LE COMPUTAZIONI E
COSTRUIRE METODI DI
"RAGIONAMENTO"
STRUTTURATO FAVORISCE
L'EXPLAINABILITY

Il problema dell'explainability

- E' molto complesso comprendere la struttura dei risultati che si ottengono dal machine learning o da tecniche di deep learning
- Quella che si definisce "Black-box"
- **E' necessario comprendere come funziona un'intelligenza artificiale?**



Deep learning e medicina

[Review](#) > [Eur Heart J](#). 2019 Jul 1;40(25):2058-2073. doi: 10.1093/eurheartj/ehz056.

Deep learning for cardiovascular medicine: a practical primer

Chayakrit Krittanawong^{1 2}, Kipp W Johnson³, Robert S Rosenson², Zhen Wang^{4 5}, Mehmet Aydar⁶, Usman Baber², James K Min⁷, W H Wilson Tang^{8 9 10}, Jonathan L Halperin², Sanjiv M Narayan¹¹

Affiliations + expand

PMID: 30815669

[Free PMC article](#)

[Review](#) > [Cancer Commun \(Lond\)](#). 2021 Nov;41(11):1100-1115. doi: 10.1002/cac2.12215.

Epub 2021 Oct 6.

Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine

Zi-Hang Chen^{1 2}, Li Lin¹, Chen-Fei Wu¹, Chao-Feng Li³, Rui-Hua Xu⁴, Ying Sun¹

Affiliations + expand

PMID: 34613667 PMCID: [PMC8626610](#) DOI: [10.1002/cac2.12215](#)

[Free PMC article](#)

Un problema di fiducia?

AI in medicine must be explainable

[Shinjini Kundu](#) ✉

[Nature Medicine](#) 27, 1328 (2021) | [Cite this article](#)

5912 Accesses | 57 Citations | 130 Altmetric | [Metrics](#)

AI algorithms used for diagnosis and prognosis must be explainable and must not rely on a black box.

In the wake of new medical devices powered by artificial intelligence and machine learning (AI/ML), a debate has erupted about whether AI's underlying logic must be intelligible. To gauge the popular sentiment, a conference posed the following question to its attendees: suppose you have cancer and need surgery to remove the tumor. Which of the two surgeons would you pick if you had to choose between a human surgeon, with a 15% chance of dying, or a robot surgeon, with a 2% chance of dying—with the caveat that no one knows how the robot operates and no questions may be asked of it? All but one of the attendees [preferred the human](#).



Perché le persone usano una tecnologia?

- Perceived value of technology (costi/benefici)
- Perceived trust of technology (affidabilità percepita)
- Explainability (comprensione)?



Come ottenere la fiducia

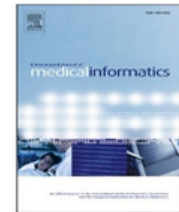
International Journal of Medical Informatics 168 (2022) 104884



Contents lists available at [ScienceDirect](#)

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf



Does AI explainability affect physicians' intention to use AI?

Chung-Feng Liu^a, Zhih-Cherng Chen^b, Szu-Chen Kuo^{c,*}, Tzu-Chi Lin^{d,*}

^a Department of Medical Research, Chi Mei Medical Center, No. 901, Zhonghua Rd., Yongkang Dist., Tainan City 710402, Taiwan

^b Division of Cardiology, Department of Internal Medicine, Chi Mei Medical Center, No. 901, Zhonghua Rd., Yongkang Dist., Tainan City 710402, Taiwan

^c Department of Pediatrics, Chi Mei Medical Center, No. 901, Zhonghua Rd., Yongkang Dist., Tainan City 710402, Taiwan

^d Department of Nursing, Chi Mei Medical Center, Liouying, No. 201, Taikang Taikang Vil., Liouying Dist., Tainan City 73657, Taiwan

ARTICLE INFO

Keywords:

Physician
Artificial intelligence (AI)
AI explainability (XAI)
Behavioral intention
Technology trust
Perceived value

ABSTRACT

Background: Artificial Intelligence (AI) is increasingly being developed to support clinical decisions for better health service quality, but the adoption of AI in hospitals is not as popular as expected. A possible reason is that the unclear AI explainability (XAI) affects the physicians' consideration of adopting the model.

Purpose: To propose and validate an innovative conceptual model aimed at exploring physicians' intention to use AI with XAI as an antecedent variable of technology trust (TT) and perceived value (PV).

Methods: A questionnaire survey was conducted to collect data from physicians of three hospitals in Taiwan. Structural equation modeling (SEM) was used to validate the proposed model and test the hypotheses.

Results: A total of 295 valid questionnaires were collected. The research results showed that physicians expressed a high intention to use AI. The XAI was found to be of great importance and had a significant impact both on AI TT and PV. We also observed that TT in AI had a significant impact on PV. Moreover, physicians' PV and TT in AI

Migliorare l'explainability

- We confirmed that TT is a significant factor that influences physicians' BI toward using AI. Therefore, medical AI developers could focus on overcoming trust issues that physicians are concerned about, especially the interpretability of ML algorithms



L'explainability è necessaria

Article 22 of the General Data Protection Regulation (GDPR) sets out the rights and obligations of the use of automated decision making. Noticeably, it introduces the right of explanation by giving individuals the right to obtain an explanation of the inference(s) automatically produced by a model, confront and challenge an associated recommendation, particularly when it might negatively affect an individual legally, financially, mentally or physically. By approving this GDPR article, the European Parliament attempted to tackle the problem related to the propagation of potentially biased inferences to society, that a computational model might have learnt from biased and unbalanced data.



Information Fusion

Volume 76, December 2021, Pages 89-106



Notions of explainability and evaluation approaches for explainable artificial intelligence

Esercitazione

*Needfinding: valutare
la necessità di
explainability e
stabilire il livello di
trust*

background

SCEGLIERE UN
PRODOTTO/SERVIZIO DIGITALE
CON O SENZA AI

strategie

GUARDARE COME I PRODOTTI
CERCANO DI AUMENTARE IL
LORO LIVELLO DI TRUST E/O
EXPLAINABILITY

needfinding

STABILIRE DEGLI OBIETTIVI E
CHIEDERE AGLI UTENTI
INFORMAZIONI SU TRUST AND
EXPLAINABILITY