L'intelligenza artificiale nel sistema delle cure: opportunità e rischi

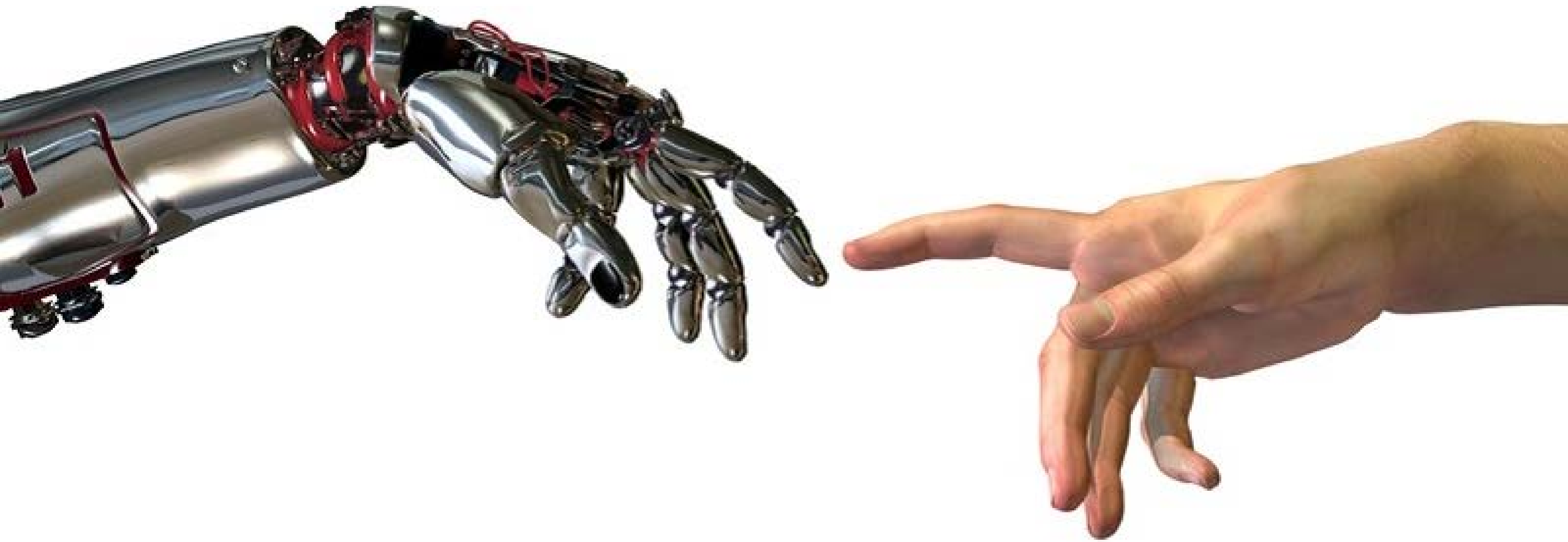
Prof. Ing. Federico Cabitza, PhD
IRCCS Istituto Ortopedico Galeazzi
Università degli Studi di Milano-Bicocca
(Dipartimento di Informatica, Sistemistica e Comunicazione)


CONFLICT OF
INTEREST

CONFLICT OF INTEREST

- ✓ Sono un ingegnere informatico (alma mater polimi);
- ✓ Sono professore di interazione uomo-macchina ad informatica;
- ✓ Insegno interazione uomo-dati (data visualization) a Data Science;
- ✓ Uso i social media per fare divulgazione su questi temi;
- ✓ Lavoro con i medici tutti i giorni...

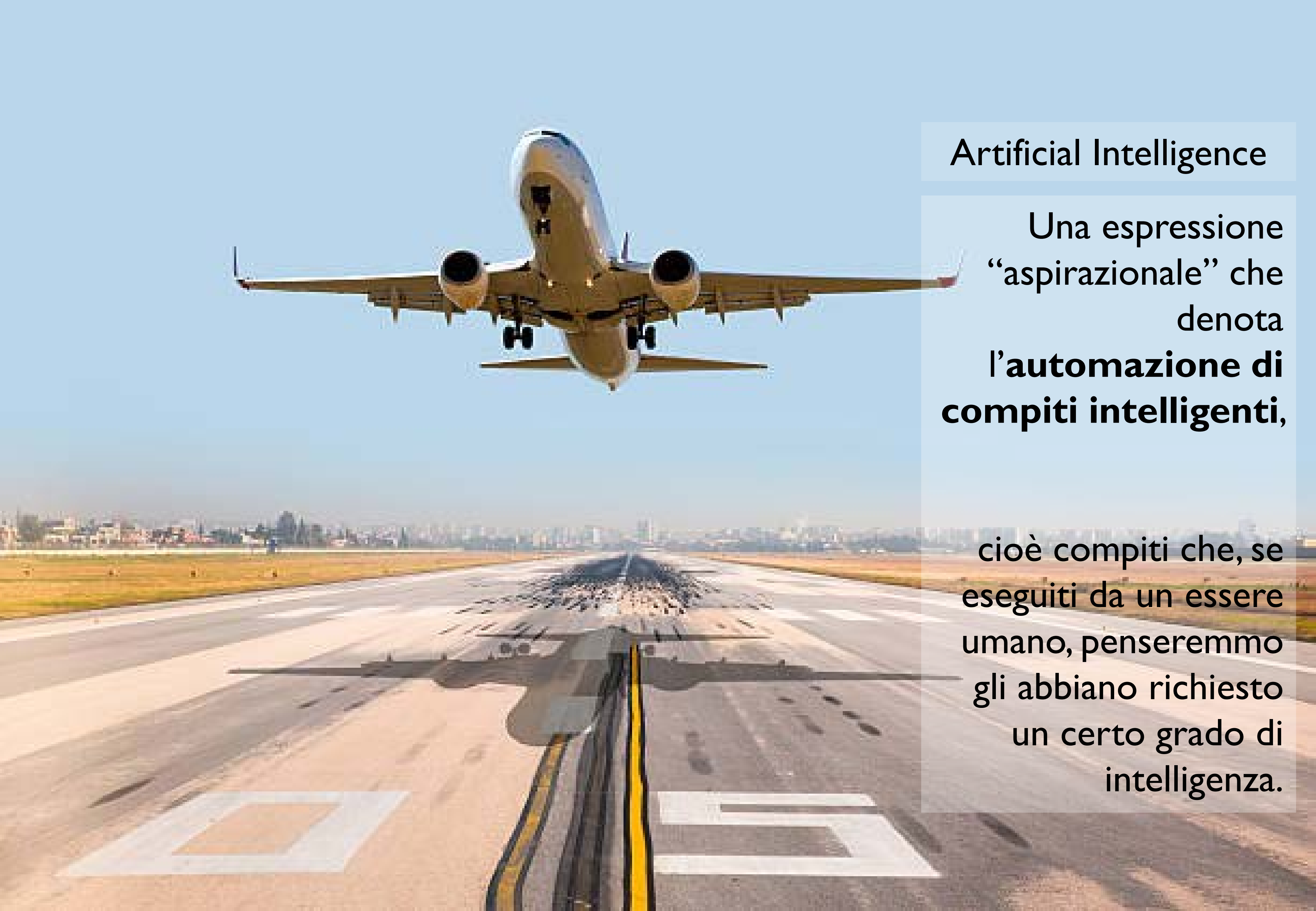
Che tipo di ingegnere informatico?



Informatica come studio e progetto della Interazione Uomo-Macchina

Interazione Uomo-Dato (interpretazione dataset)

Interazione Uomo-AI (interpr. supporto decisionale)



Artificial Intelligence

Una espressione “aspirazionale” che denota l’**automazione di compiti intelligenti**,

cioè compiti che, se eseguiti da un essere umano, penseremmo gli abbiano richiesto un certo grado di intelligenza.

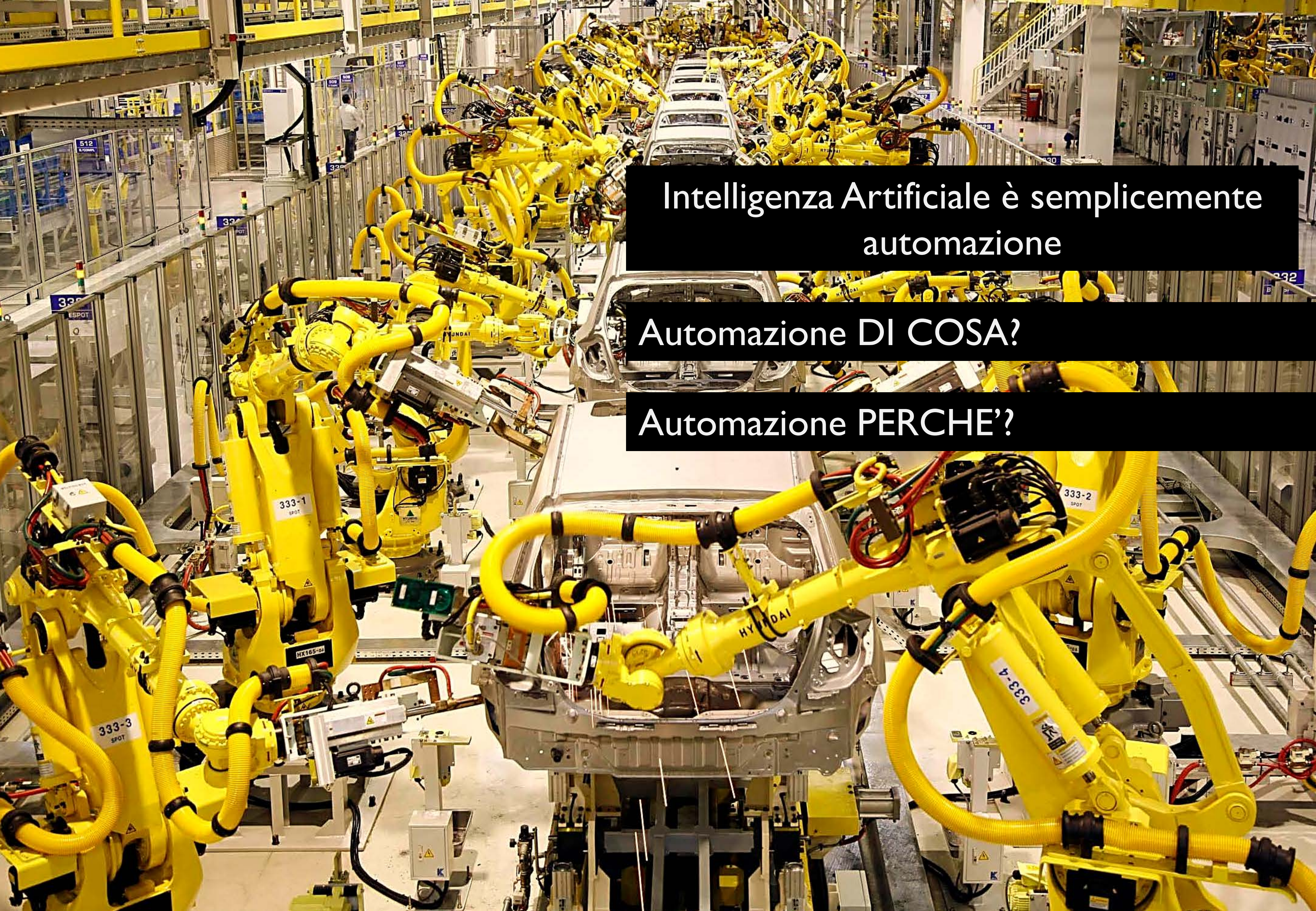
Artificial Intelligence

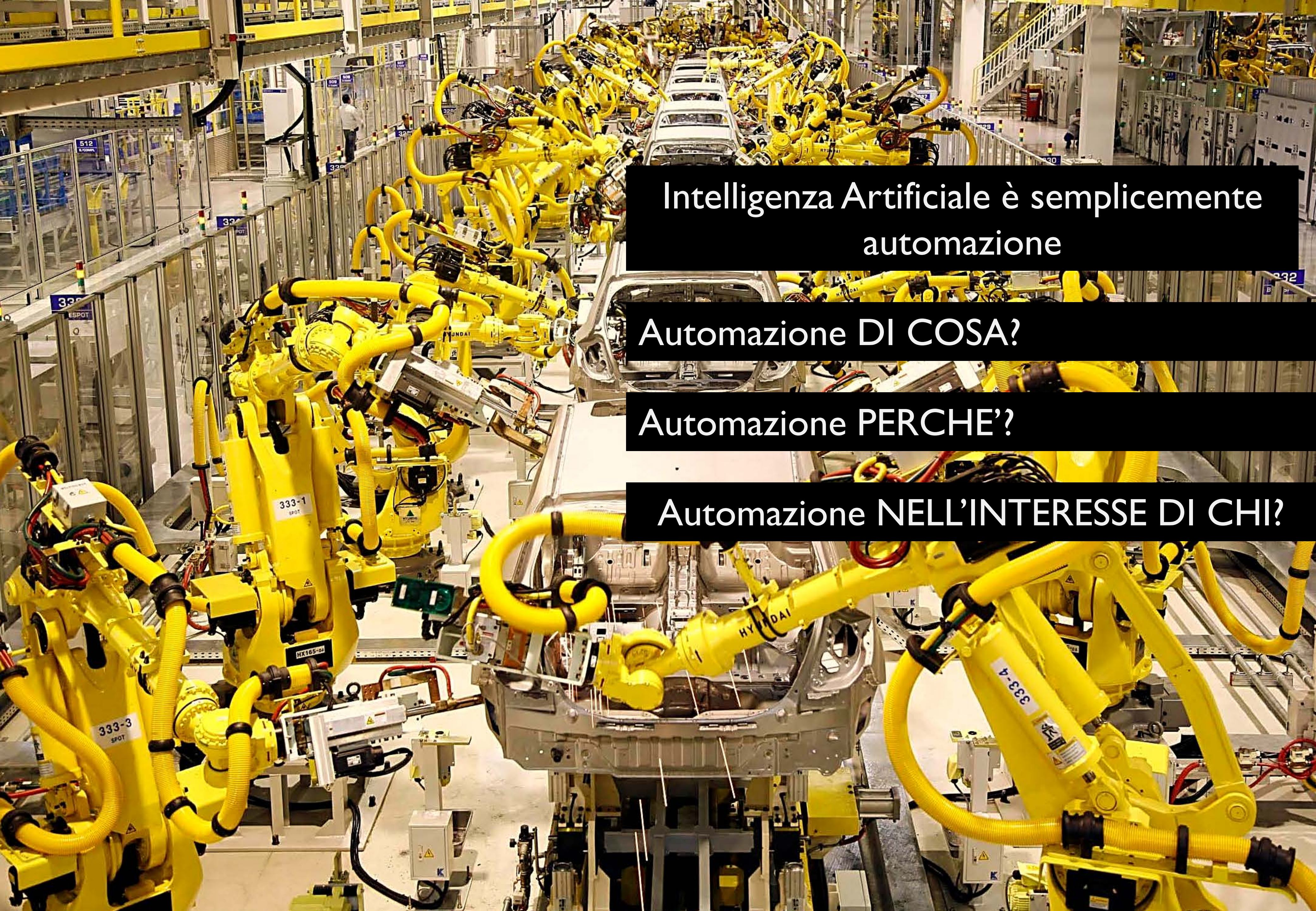
L’area professionale a cavallo di logica formale, ingegneria e informatica in cui operano e lavorano quelli che costruiscono sistemi AI.

Intelligenza Artificiale è semplicemente automazione

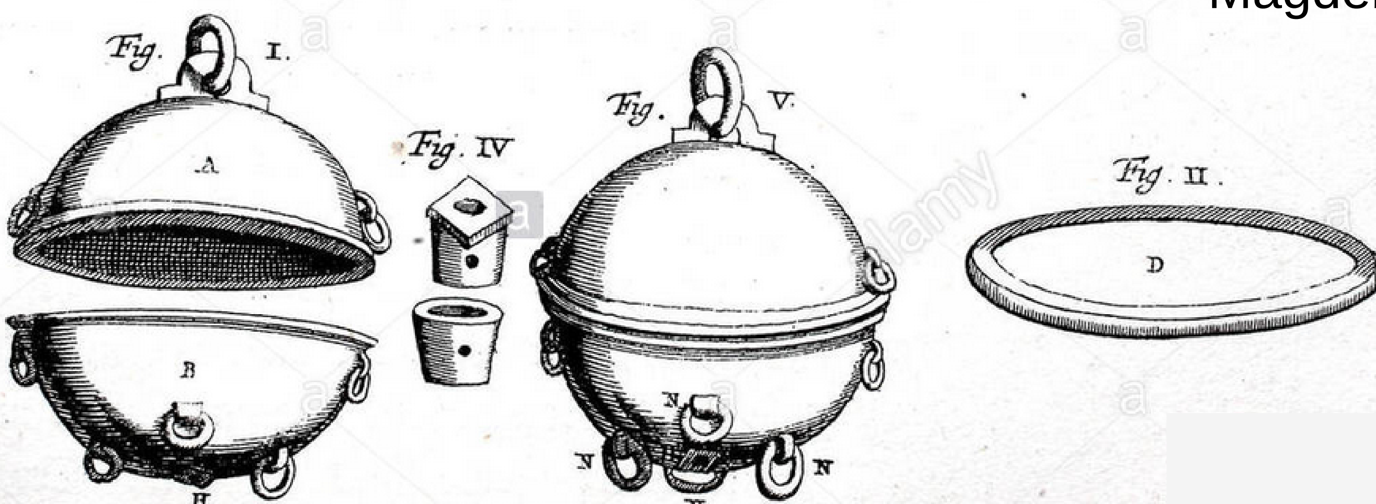
Intelligenza Artificiale è semplicemente automazione

Automazione DI COSA?


Intelligenza Artificiale è semplicemente automazione
Automazione DI COSA?
Automazione PERCHE'?

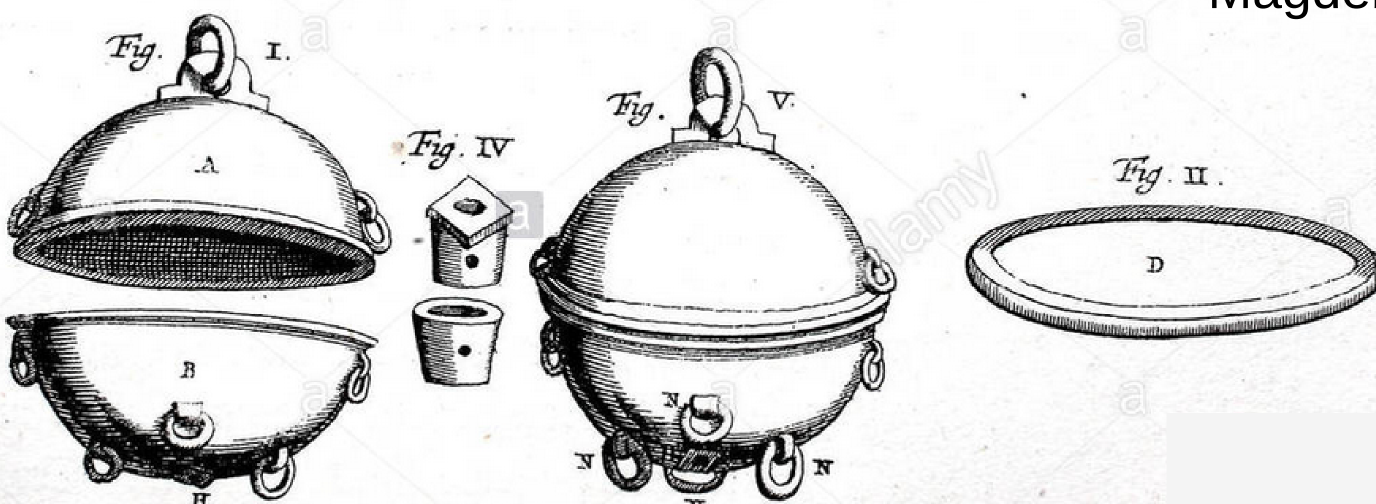

Intelligenza Artificiale è semplicemente automazione
Automazione DI COSA?
Automazione PERCHE'?
Automazione NELL'INTERESSE DI CHI?

Magdeburg 1654



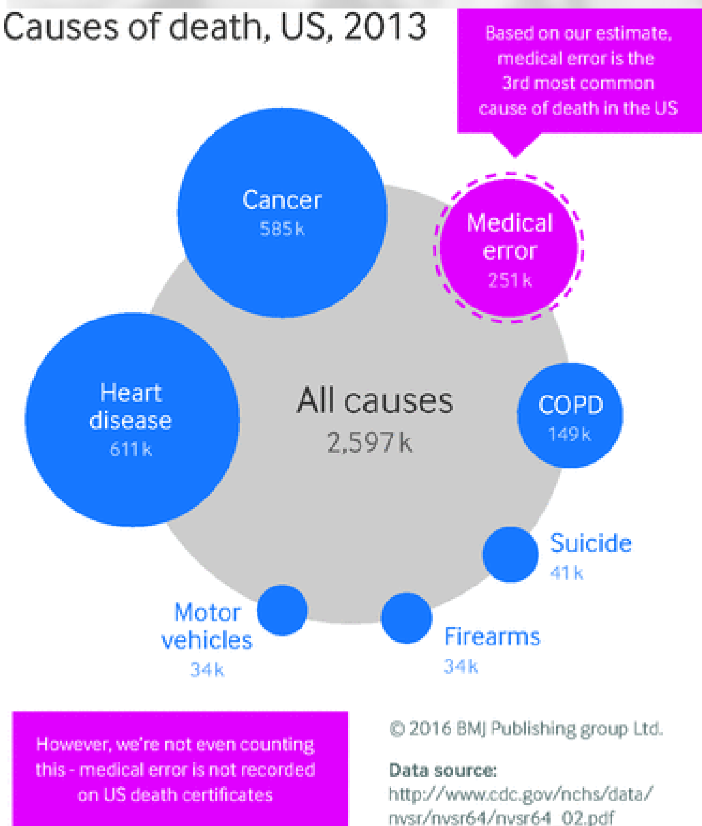
EFFICACIA EFFICIENZA

Magdeburg 1654



Efficacia

Efficienza



thebmj

BMJ 2016;353:i2139 doi: 10.1136/bmj.i2139 (Published 3 May 2016)

Page 1 of 5

ANALYSIS

Medical error—the third leading cause of death in the US

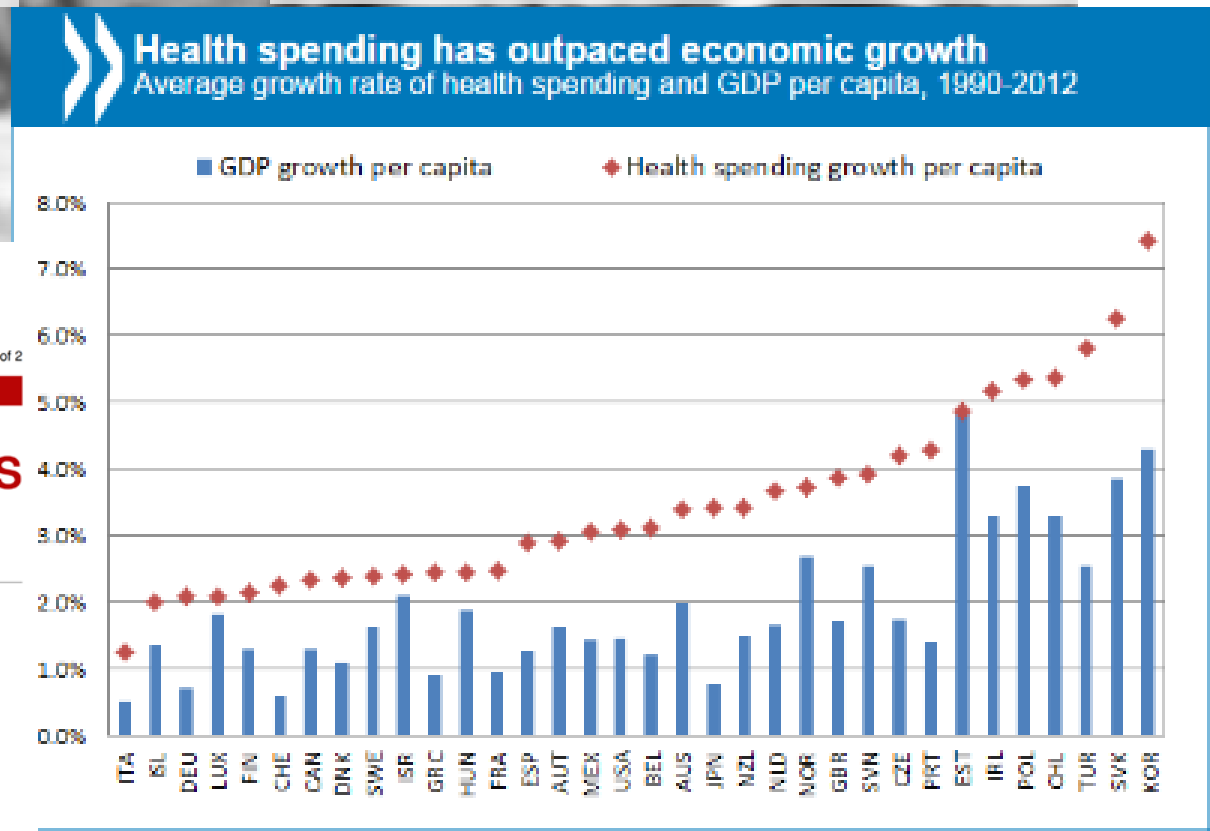
Medical error is not included on death certificates or in rankings of cause of death. **Martin Makary** and **Michael Daniel** assess its contribution to mortality and call for better reporting

Martin A Makary *professor*, Michael Daniel *research fellow*

Department of Surgery, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA

EFFICACIA

EFFICIENZA



thebmj

BMJ 2017;358:j3360 doi: 10.1136/bmj.j3360 (Published 14 July 2017)

EDITORIALS

Burnout among doctors

A system level problem requiring a system level response

Jane B Lemaire *clinical professor*[1], Jean E Wallace *professor*[2]

[1]Cumming School of Medicine, University of Calgary, Alberta, Canada; [2]Department of Sociology, Faculty of Arts, University of Calgary

Although doctors have a professional responsibility to be at

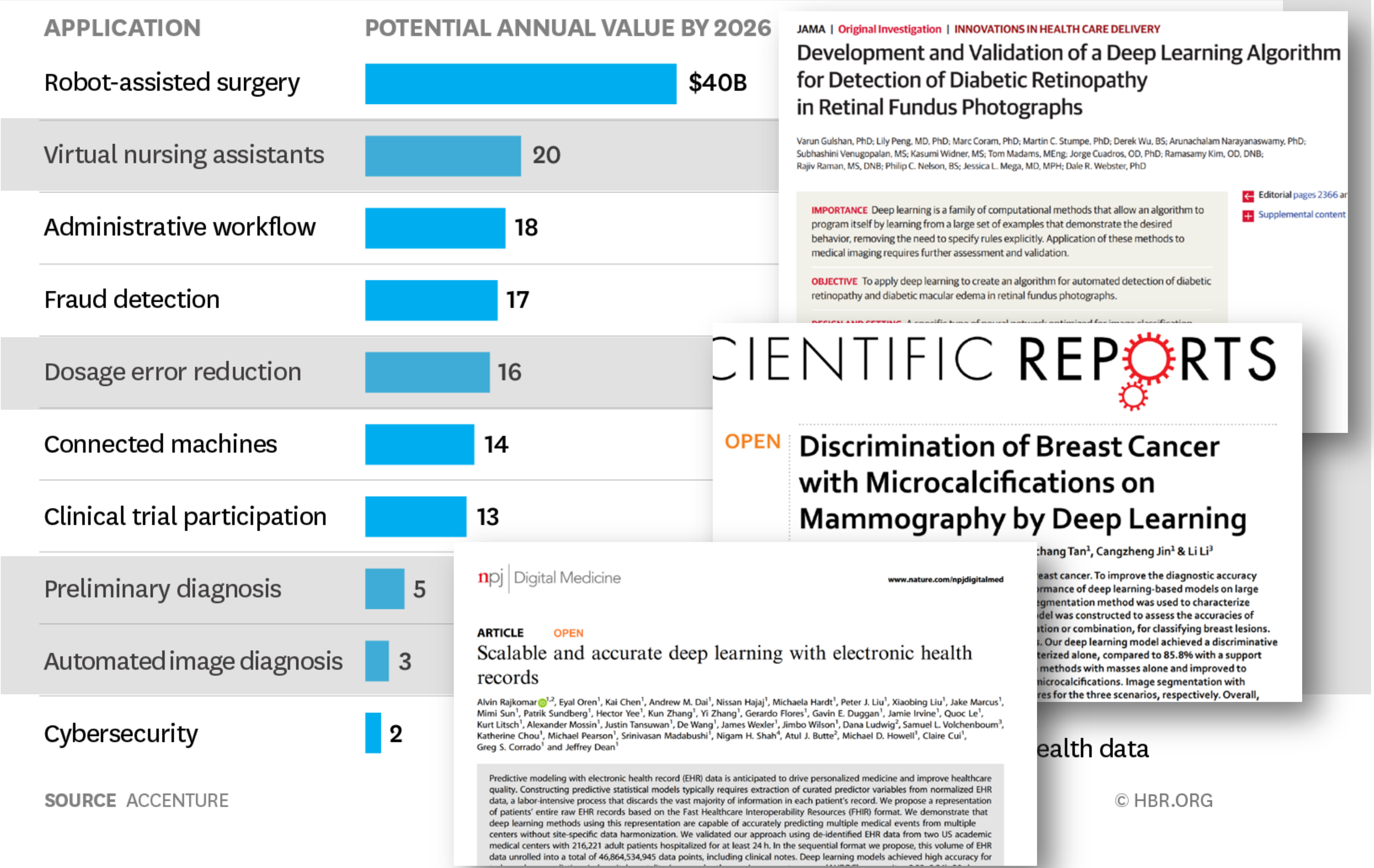
disorganised rotations and inadequate supervision are also

10 AI Applications That Could Change Health Care

APPLICATION	POTENTIAL ANNUAL VALUE BY 2026	KEY DRIVERS FOR ADOPTION
Robot-assisted surgery	$40B	Technological advances in robotic solutions for more types of surgery
Virtual nursing assistants	20	Increasing pressure caused by medical labor shortage
Administrative workflow	18	Easier integration with existing technology infrastructure
Fraud detection	17	Need to address increasingly complex service and payment fraud attempts
Dosage error reduction	16	Prevalence of medical errors, which leads to tangible penalties
Connected machines	14	Proliferation of connected machines/devices
Clinical trial participation	13	Patent cliff; plethora of data; outcomes-driven approach
Preliminary diagnosis	5	Interoperability/data architecture to enhance accuracy
Automated image diagnosis	3	Storage capacity; greater trust in AI technology
Cybersecurity	2	Increase in breaches; pressure to protect health data

SOURCE ACCENTURE

Analisi e Interpretazione dei segni, sintomi, immagini e dati di laboratorio per formulare diagnosi e raccomandazioni



https://hbr.org/2018/05/10-promising-ai-applications-in-health-care?platform=hootsuite

Decision support systems

Decision support systems

Levels of Automation of Decision and Action Selection

HIGH	10. The computer decides everything, acts autonomously, ignoring the human.
	9. informs the human only if it, the computer, decides to
	8. informs the human only if asked, or
	7. executes automatically, then necessarily informs the human, and
	6. allows the human a restricted time to veto before automatic execution, or
	5. executes that suggestion if the human approves, or
	4. suggests one alternative
	3. narrows the selection down to a few, or
	2. The computer offers a complete set of decision/action alternatives, or
LOW	1. The computer offers no assistance: human must take all decisions and actions.

- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, *30*(3), 286-297.

DECISION SUPPORT SYSTEMS

HCI

LEVELS OF AUTOMATION OF DECISION AND ACTION SELECTION

HIGH	10. The computer decides everything, acts autonomously, ignoring the human.
	9. informs the human only if it, the computer, decides to
	8. informs the human only if asked, or
	7. executes automatically, then necessarily informs the human, and
	6. allows the human a restricted time to veto before automatic execution, or
	5. executes that suggestion if the human approves, or
	4. suggests one alternative
	3. narrows the selection down to a few, or
	2. The computer offers a complete set of decision/action alternatives, or
LOW	1. The computer offers no assistance: human must take all decisions and actions.

- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, *30*(3), 286-297.

Decision support systems

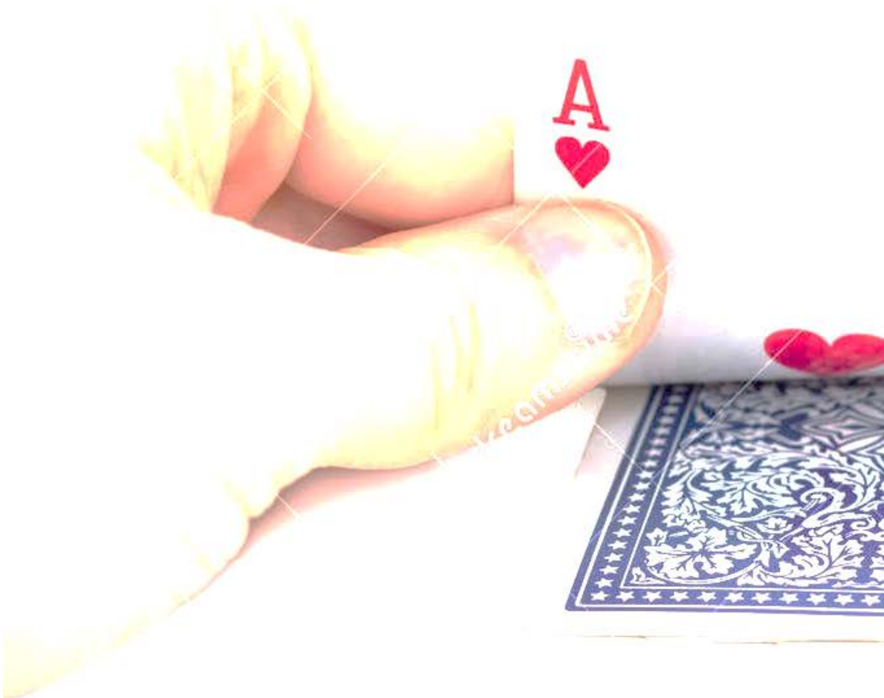
Levels of Automation of Decision and Action Selection

HIGH	10. The computer decides everything, acts autonomously, ignoring the human.
	9. informs the human only if it, the computer, decides to
	8. informs the human only if asked, or
	7. executes automatically, then necessarily informs the human, and
	6. allows the human a restricted time to veto before automatic execution, or
	5. executes that suggestion if the human approves, or
	4. suggests one alternative
	3. narrows the selection down to a few, or
	2. The computer offers a complete set of decision/action alternatives, or
LOW	1. The computer offers no assistance: human must take all decisions and actions.

- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans*, *30*(3), 286-297.

Predizione dell'esito di un intervento chirurgico

Predizione dell’esito di un intervento chirurgico



Hip:: Mental score :: 3 months

Prediction of the mental score from the SF12 questionnaire (MCS) of hip surgery (first intervention) after 3 months.

Selected features:

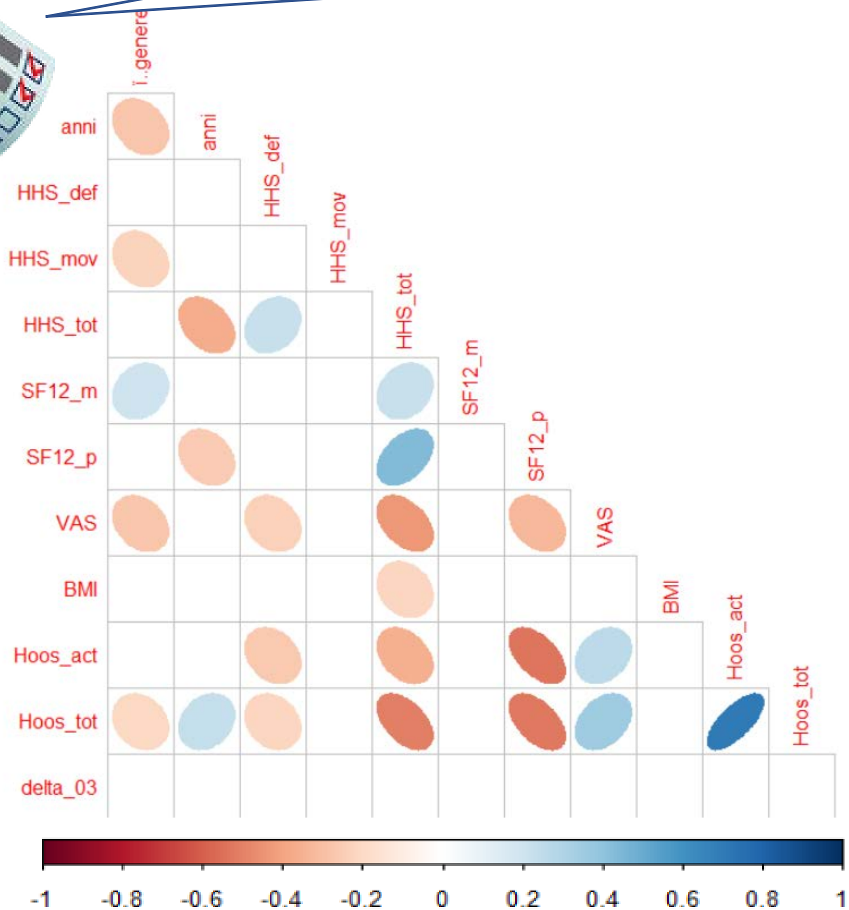
1. Age,
2. gender,
3. pre-op HHS deformity,
4. per-op HSS movement,
5. per-op HSS total
6. pre-op SF12 mental score,
7. pre-op SF12 physical score.
8. Pre-op pain VAS
9. Pre-op BMI
10. Preop Hoos-ps

Target variable:

- Delta SF12 mental score (MCS) 0-3 months

Predizione dell'esito di un intervento chirurgico

Nessun imbroglio: le variabil di input non hanno alcuna correlazione con l'output.



Hip:: Mental score :: 3 months

Prediction of the mental score from the SF12 questionnaire (MCS) of hip surgery (first intervention) after 3 months.

Selected features:

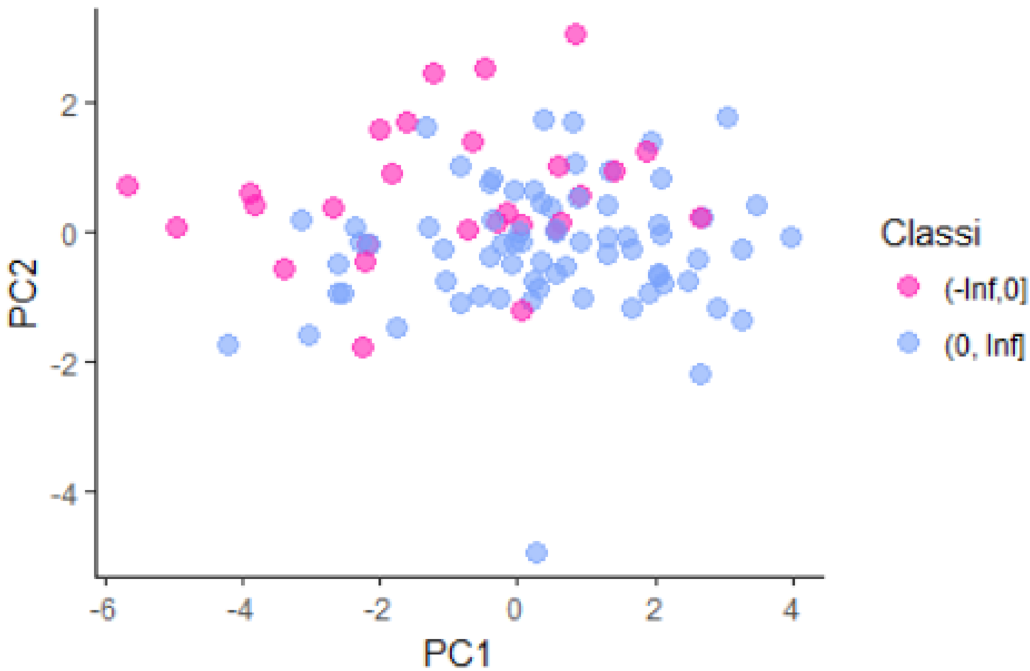
1. Age,
2. gender,
3. pre-op HHS deformity,
4. per-op HSS movement,
5. per-op HSS total
6. pre-op SF12 mental score,
7. pre-op SF12 physical score.
8. Pre-op pain VAS
9. Pre-op BMI
10. Preop Hoos-ps activity

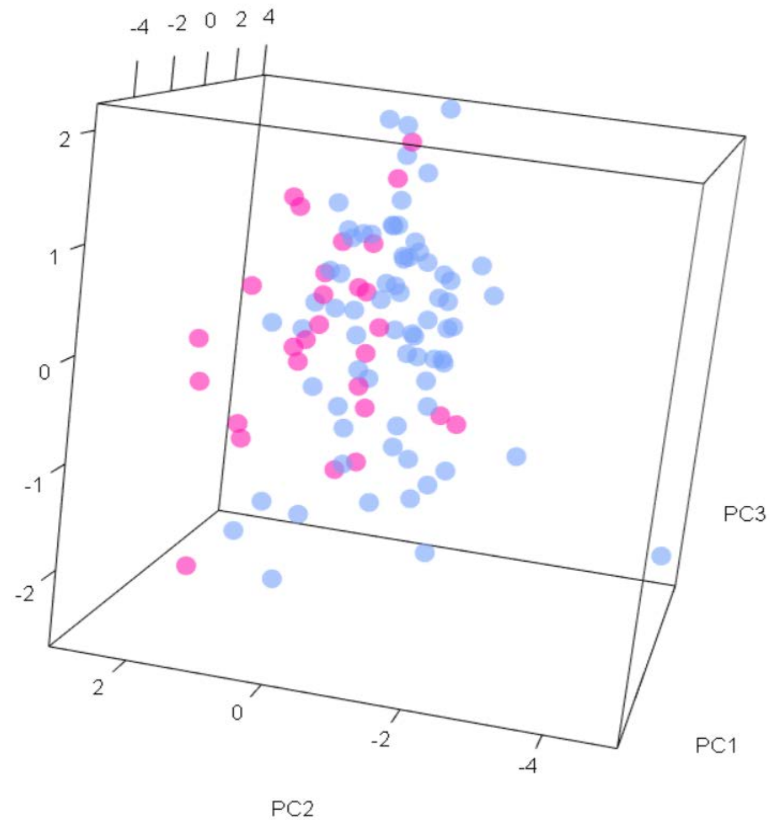
Target variable:

- Delta SF12 mental score (MCS) 0-3 months

Predizione dell'esito di un intervento chirurgico

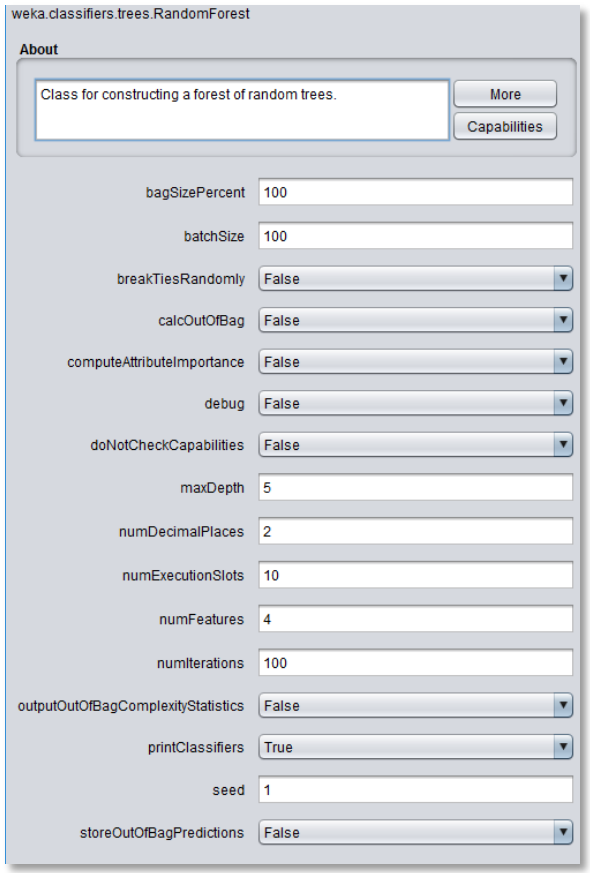
Nessun imbroglio: peggiorati e migliorati sono effettivamente "mischiati".





Predizione dell'esito di un intervento chirurgico

Il modello più accurato per questo problema è risultato il Random Forest.



Hip:: Mental score :: 3 months

Prediction of the mental score from the SF12 questionnaire (MCS) of hip surgery (first intervention) after 3 months.

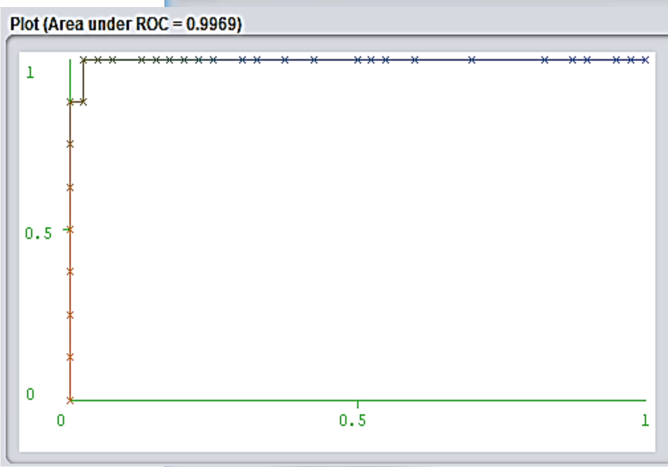
Selected features:

1. Age,
2. gender,
3. pre-op HHS deformity,
4. per-op HSS movement,
5. per-op HSS total
6. pre-op SF12 mental score,
7. pre-op SF12 physical score.
8. Pre-op pain VAS
9. Pre-op BMI
10. Preop Hoos-ps activity

Target variable:

- Delta SF12 mental score (MCS) 0-3 months

PREDIZIONE DELL'ESITO DI UN INTERVENTO CHIRURGICO



Accuracy: 97.9%
TP rate: 97.4%
FP rate: 0.04%
Precision: 98.1%
Recall: 97.9%
F-score: 98.0% (on test set)

HIP:: MENTAL SCORE :: 3 MONTHS

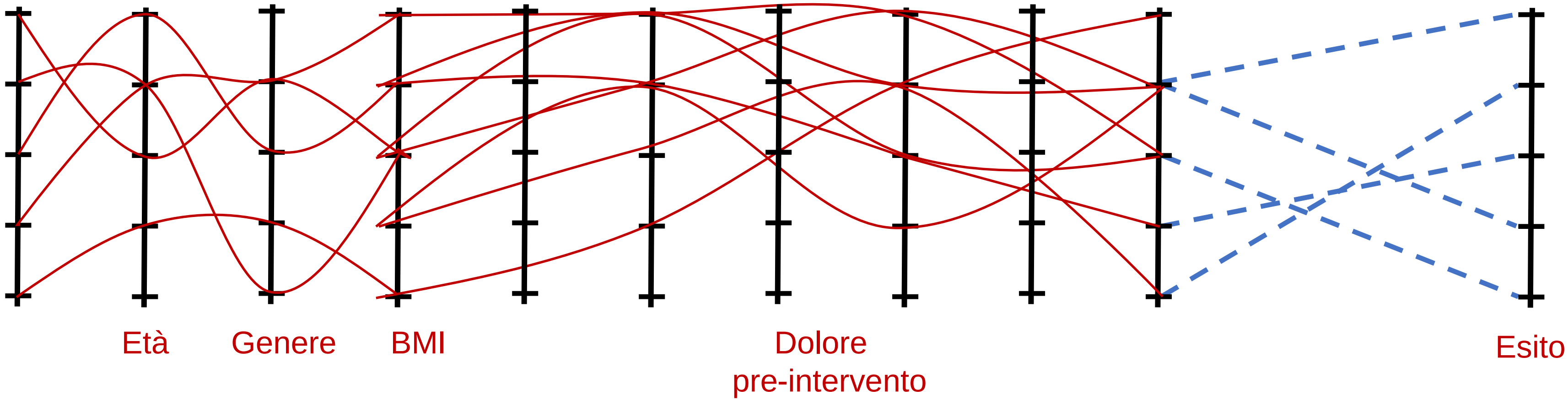
Prediction of the mental score from the SF12 questionnaire (MCS) of hip surgery (first intervention) after 3 months.

Selected features:
1. Age,
2. gender,
3. pre-op HHS deformity,
4. ...-op H...
5. ...op ... total
6. pre-op SF12 mental score,
7. pre-op SF12 physical score.
8. Pre-op pain VAS
9. ...re-o... BMI
10. ...reo... loos-... ...ctiv...

...get variable:
- Delta SF12 mental score (MCS) 0-3 months

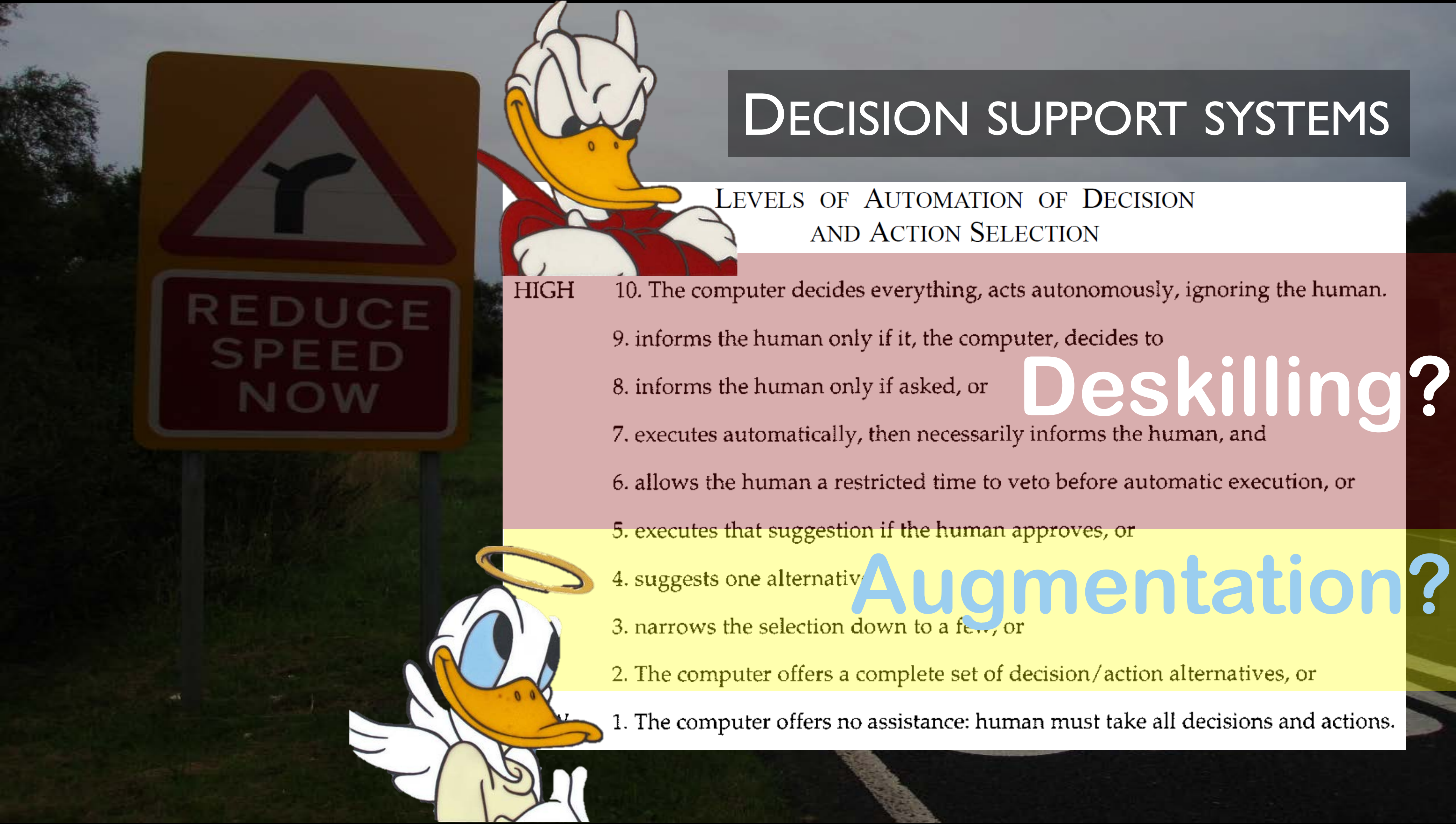
99%

A partire da solo 10 variabili iniziali



99%

Decision support systems



LEVELS OF AUTOMATION OF DECISION AND ACTION SELECTION

HIGH 10. The computer decides everything, acts autonomously, ignoring the human.

9. informs the human only if it, the computer, decides to

8. informs the human only if asked, or

7. executes automatically, then necessarily informs the human, and

6. allows the human a restricted time to veto before automatic execution, or

5. executes that suggestion if the human approves, or

4. suggests one alternativ

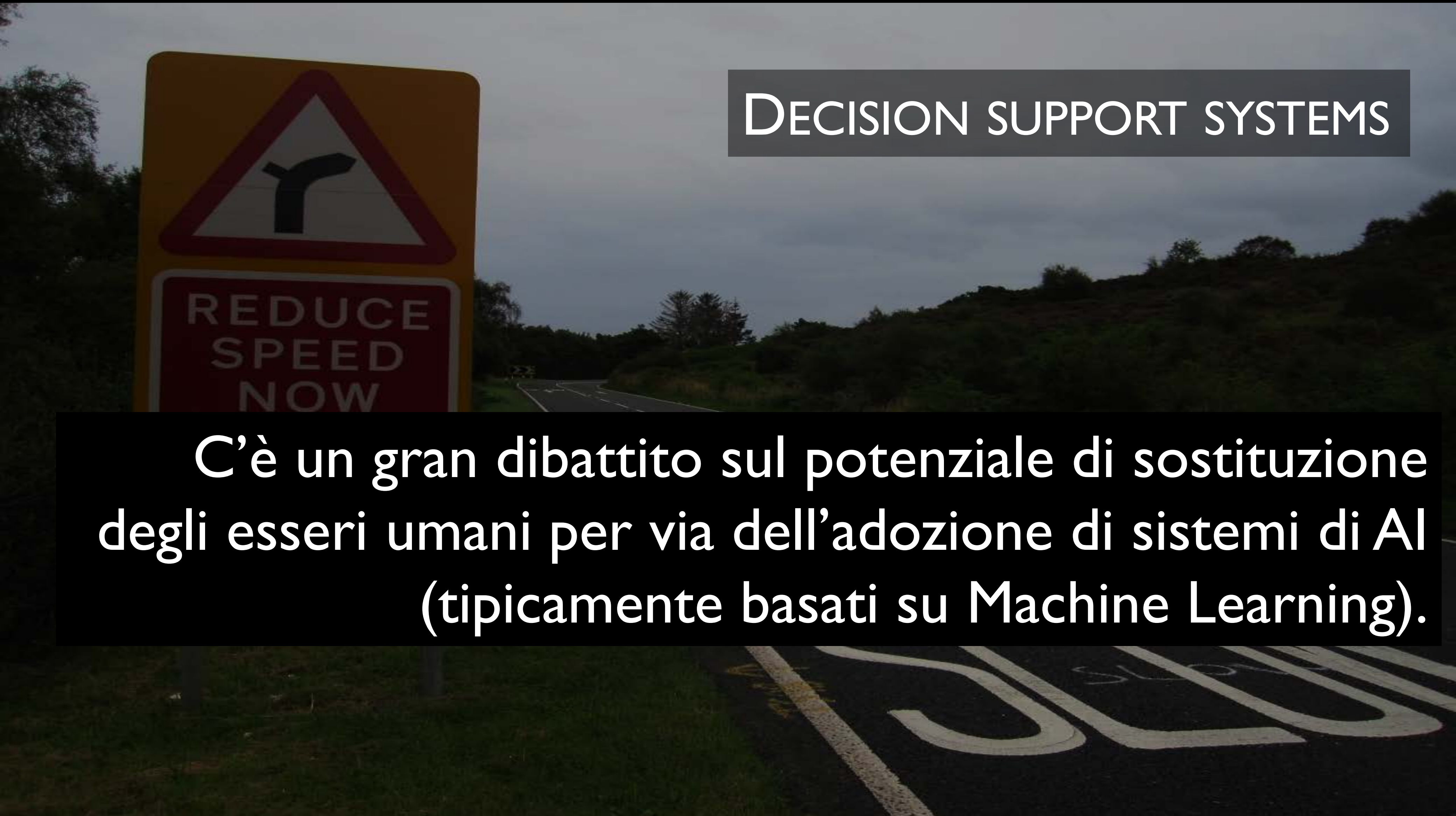
3. narrows the selection down to a fe , or

2. The computer offers a complete set of decision/action alternatives, or

1. The computer offers no assistance: human must take all decisions and actions.

Decision support systems

C’è un gran dibattito sul potenziale di sostituzione degli esseri umani per via dell’adozione di sistemi di AI (tipicamente basati su Machine Learning).

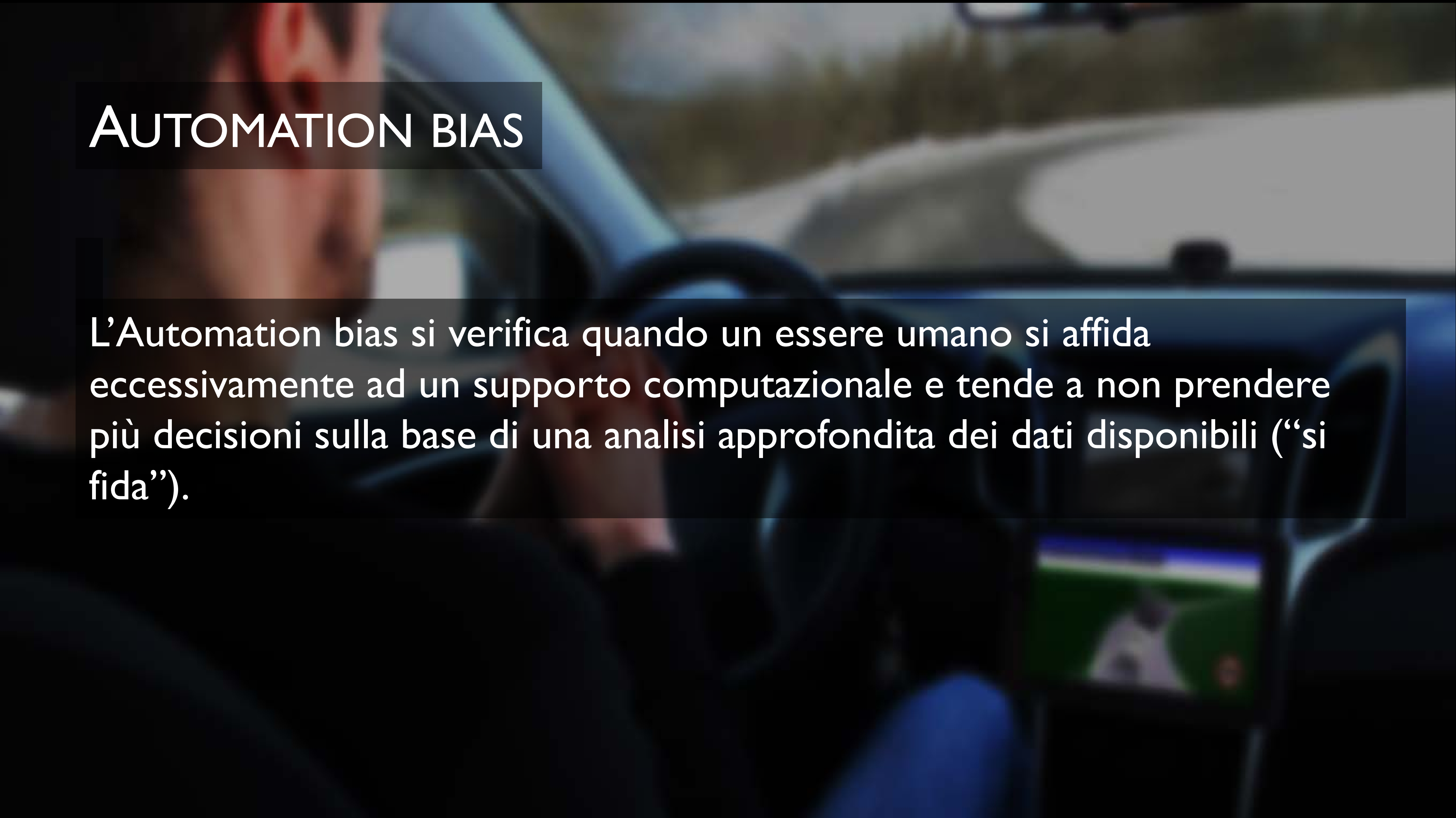


DECISION SUPPORT SYSTEMS

C’è un gran dibattito sul potenziale di sostituzione degli esseri umani per via dell’adozione di sistemi di AI (tipicamente basati su Machine Learning).

Ma forse è meglio capire che competenze perdiamo se ci affidiamo troppo a questi strumenti.

Automation bias



Automation bias

L'Automation bias si verifica quando un essere umano si affida eccessivamente ad un supporto computazionale e tende a non prendere più decisioni sulla base di una analisi approfondita dei dati disponibili ("si fida").

- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human Factors: The Journal of the Human Factors and Ergonomics Society, 52(3), 381-410.

Recenti Prog Med 2017; 108: 397-401

Potenziali conseguenze inattese dell'uso di sistemi di intelligenza artificiale oracolari in medicina

FEDERICO CABITZA[1,2*], CAMILLA ALDERIGHI[3*], RAFFAELE RASOINI[3,4*], GIAN FRANCO GENSINI[3*]

[1]IRCCS Istituto Ortopedico Galeazzi, Milano; [2] Dipartimento di Informatica, Sistemistica e Comunicazione, Università di Milano-Bicocca, Milano; [3]CESMAV - Centro Studi Medicina Avanzata, Firenze; [4]IFCA Istituto Fiorentino di Cura e Assistenza, Firenze.
*Florence EBM/Renaissance Group.

Pervenuto il 16 agosto 2017. Accettato dopo revisione il 31 agosto 2017.

Riassunto. I sistemi di supporto decisionale basati sul *machine learning* (ML) in medicina stanno raccogliendo un crescente interesse grazie a recenti pubblicazioni che ne hanno evidenziato l'elevata accuratezza diagnostica in specifici contesti clinici. Tuttavia, agli ipotetici vantaggi derivanti dall'applicazione dei sistemi di intelligenza artificiale in campo medico, vanno criticamente affiancati alcuni potenziali inconvenienti. Alla luce dell'attuale mancanza di studi sugli effetti collaterali dell'applicazione di questi nuovi supporti decisionali nella pratica medica,

"Handle with care": about the potential unintended consequences of oracular artificial intelligence systems in medicine.

Summary. Decisional support systems based on *machine learning* (ML) in medicine are gaining a growing interest as some recent articles have highlighted the high diagnostic accuracy exhibited by these systems in specific medical contexts. However, it is implausible that any potential advantage can be obtained without some potential

Recenti Prog Med 2017; 108: 397-401

JAMA® The Journal of the American Medical Association

VIEWPOINT

Unintended Consequences of Machine Learning in Medicine

Federico Cabitza, PhD
Department of Informatics, University of Milano-Bicocca, Milan, Italy; and IRCCS Istituto Ortopedico Galeazzi, Milan, Italy.

Raffaele Rasoini, MD
Centro Studi Medicina Avanzata, Florence, Italy.

Gian Franco Gensini, MD
Centro Studi Medicina Avanzata, Florence, Italy.

Over the past decade, machine learning techniques have made substantial advances in many domains. In health care, global interest in the potential of machine learning has increased; for example, a deep learning algorithm has shown high accuracy in detecting diabetic retinopathy.[1] There have been suggestions that machine learning will drive changes in health care within a few years, specifically in medical disciplines that require more accurate prognostic models (eg, oncology) and those based on pattern recognition (eg, radiology and pathology).

However, comparative studies on the effectiveness of machine learning–based decision support systems (ML-DSS) in medicine are lacking, especially regarding the effects on health outcomes. Moreover, the introduction of new technologies in health care has not always been straightforward or without unintended and

the expense of other elements that are more difficult or impossible to easily describe. Relying on ML-DSS requires considering digital data as reliable and complete representations of the phenomena that these data are supposed to render in a discrete and trustworthy form. This may be a problem when the clinical context is not represented, particularly if physicians lose awareness of the existence of clinical elements that are not included in the clinical record.

Such lack of information may lead to partial or misleading interpretations of ML-DSS diagnostics and therapeutic or prognostic outputs. It also could lead to reduced interest in and decreased ability to perform holistic evaluations of patients, with loss of valuable and irreducible aspects of the human experience such as psychological, relational, social, and organizational issues.

Recenti Prog Med 2017; 108: 397-401

JAMA® The Journal of the American Medical Association

Opinion

VIEWPOINT

Unintended Consequences of Machine Learning in Medicine

Federico Cabitza, PhD
Department of Informatics, University of Milano-Bicocca, Milan, Italy; and IRCCS Istituto Ortopedico Galeazzi, Milan, Italy.

Raffaele Rasoini, MD
Centro Studi Medicina Avanzata, Florence, Italy.

Gian Franco Gensini, MD
Centro Studi Medicina Avanzata, Florence, Italy.

Over the past decade, machine learning techniques have made substantial advances in many domains. In health care, global interest in the potential of machine learning has increased; for example, a deep learning algorithm has shown high accuracy in detecting diabetic retinopathy.[1] There have been suggestions that machine learning will drive changes in health care within a few years, specifically in medical disciplines that require more accurate prognostic models (eg, oncology) and those based on pattern recognition (eg, radiology and pathology).

However, comparative studies on the effectiveness of machine learning–based decision support systems (ML-DSS) in medicine are lacking, especially regarding the effects on health outcomes. Moreover, the introduction of new technologies in health care has not always been straightforward or without unintended and

the expense of other elements that are more difficult or impossible to easily describe. Relying on ML-DSS requires considering digital data as reliable and complete representations of the phenomena that these data are supposed to render in a discrete and trustworthy form. This may be a problem when the clinical context is not represented, particularly if physicians lose awareness of the existence of clinical elements that are not included in the clinical record.

Such lack of information may lead to partial or misleading interpretations of ML-DSS diagnostics and therapeutic or prognostic outputs. It also could lead to reduced interest in and decreased ability to perform holistic evaluations of patients, with loss of valuable and irreducible aspects of the human experience such as psychological, relational, social, and organizational issues

397 | Pro

JAMA®
The Journal of the
American Medical
Association

Dovrebbe poi divenire ben radicata, tra gli informatici e tra gli esperti di ML, l'esigenza di non trascurare l'incertezza insita nell'interpretazione di ogni fenomeno in medicina e anzi prioritario considerare tale incertezza, e valorizzarla, nell'ambito degli algoritmi di definizione dei modelli predittivi e nella interpretazione dei loro risultati[17].

Andrebbe inoltre evidenziato e studiato il pericolo di sovra-affidamento e di eccessiva dipendenza da sistemi di accuratezza che definiamo "oracolare", cioè molto elevata ma non associata a spiegazioni esplicite e significative per gli operatori coinvolti[8], con gli effetti secondari di deskilling e desensibilizzazione al contesto clinico.

AUTOMATION BIAS

Automatic
thorough a
automatica

Mosier and
using the c
vigilant ir

"Automation bias è stato osservato sia nell'inesperto che nell'esperto, non può essere evitato con una adeguata formazione o insieme di istruzioni,e può condizionare la decisione, inducendo errori sia di omissione che di commissione (overdiagnosis), da parte di individui come di team." *

- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. Human Factors: The Journal of the Human Factors and Ergonomics Society, 52(3), 381-410.

Automation bias

Automation bias

OverReliance

OverConfidence

Overdependence

Automation bias

OverReliance

OverConfidence

Overdependence

Desensitivity to context

Automation bias

OverReliance

OverConfidence

Overdependence

Desensitivity to context

Deskilling

Automation bias

OverReliance

OverConfidence

Overdependence

Desensitivity to context

deskilling

Automation bias

OverReliance

OverConfidence | Overdependence

Desensitivity to context

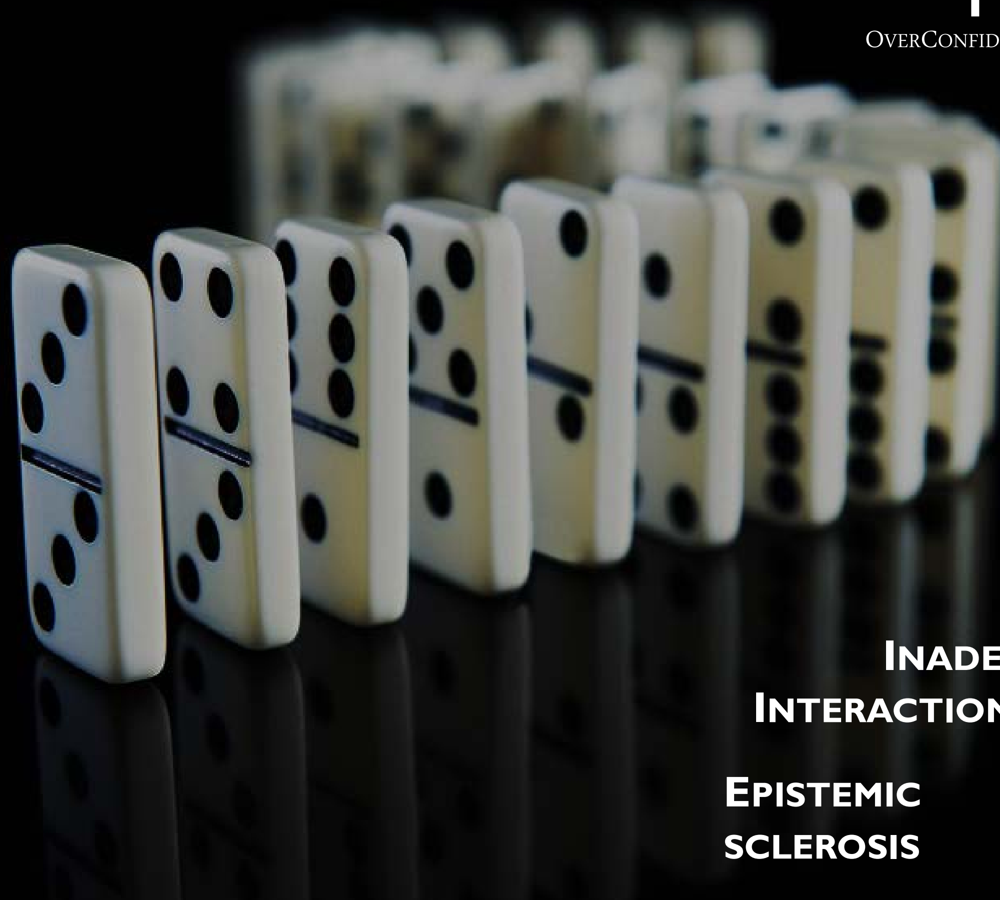
Deskilling

Overdiagnosis

Overtest

Overtreatment

AUTOMATION BIAS



OVERRELIANCE

OVERCONFIDENCE

OVERDEPENDENCE

DESENSITIVITY TO CONTEXT

DESKILLING

INADEQUATE INTERACTION W/ AI

EPISTEMIC SCLEROSIS

OVERDIAGNOSIS

OVERTEST

OVERTREATMENT

Gli algoritmi AI sono come farmaci?

ALGORITHMS ARE THE NEW DRUGS?
REFLECTIONS FOR A CULTURE OF IMPACT ASSESSMENT AND VIGILANCE

Hugh Harvey
Kheiron Medical Technologies
40 islington High street, London, N1 8EQ, UK

Federico Cabitza
Università degli Studi di Milano-Bicocca
IRCCS Istituto Ortopedico Galeazzi
Viale Sarca 336, 20126 Milano, Italy

ABSTRACT

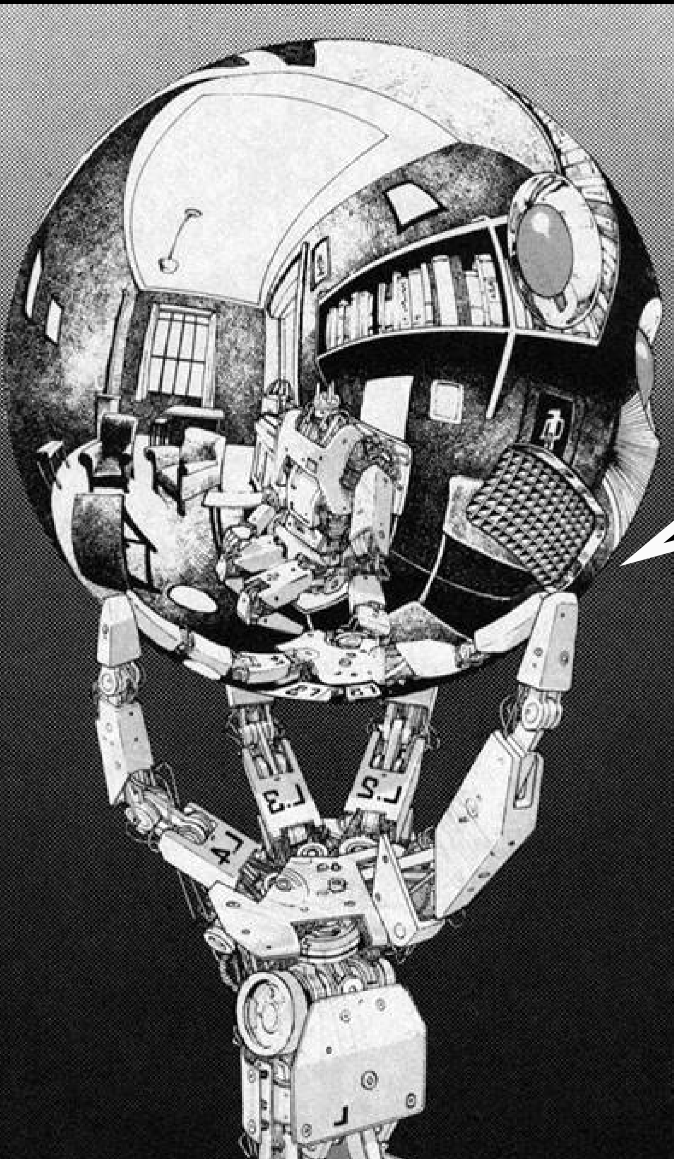
In this paper we propose that Artificially Intelligent (AI) systems in medicine are *the new drugs*. Through study and comparison of the existing corporate pharmaceutical models, we advocate a new approach in how the performance of AI solutions (both for diagnostic and prognostic aims) is measured and promoted, on how the impact of these solutions is assessed and evaluated in terms of both outcome and safety (about which the new term *technovigilance* has been introduced), and how education and a better awareness of the impact of these systems on heath care by multiple stakeholders should be promoted. Any drug must be evaluated in terms of efficacy and safety, but also in terms of real-world effectiveness, efficiency and cost-effectiveness, that is in terms of its role as a component of a wider and more complex socio-technical system where it must fit a broader

Gli algoritmi AI sono come farmaci?

Sì!

ALGORITHMS ARE THE NEW DRUGS?
REFLECTIONS FOR A CULTURE OF IMPACT ASSESSMENT AND VIGILANCE

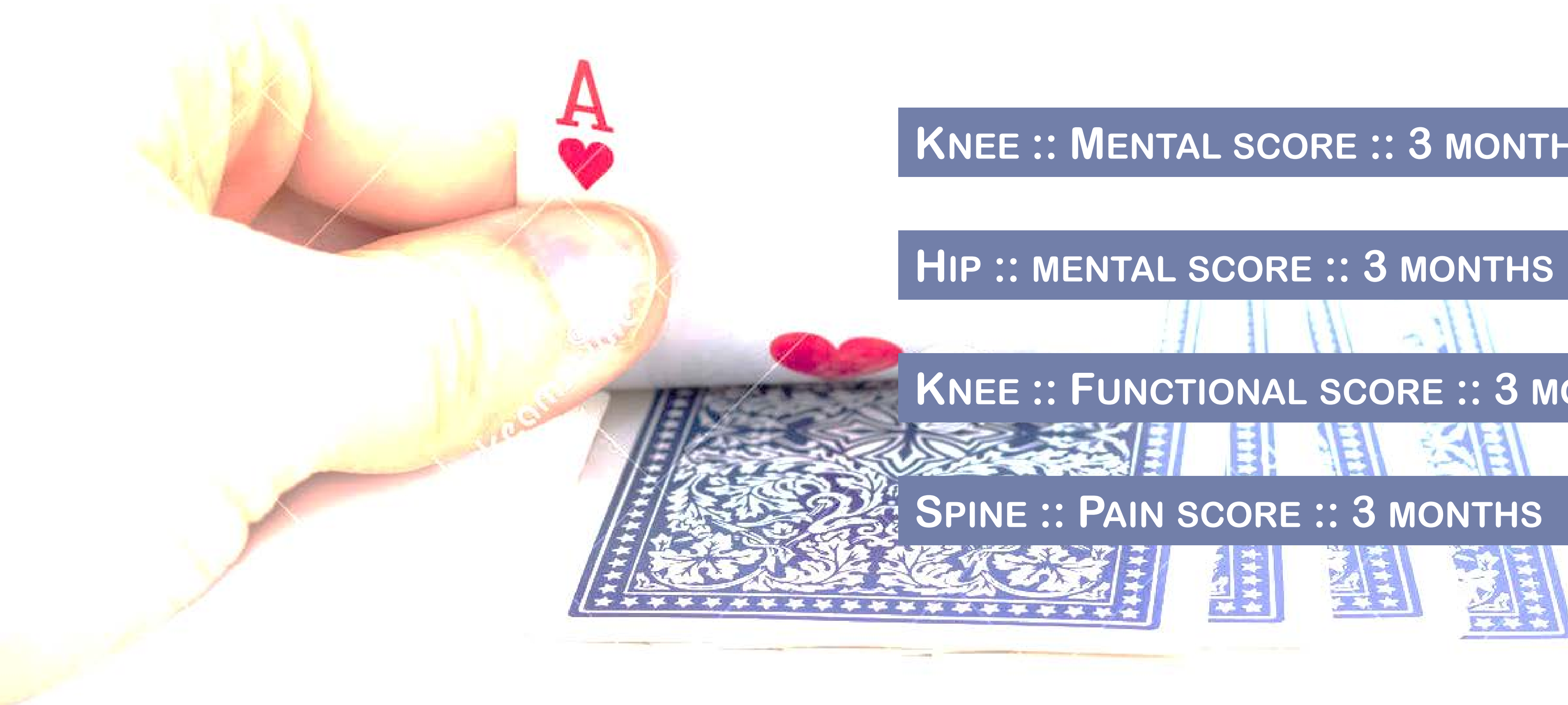
- ✓ Sì, perché la loro adozione può apportare un potenziale di benefici.
- ✓ Sì, perché questo potenziale DEVE essere valutato in condizioni real-world (confrontando end point primari e secondari in bracci sperimentali omogenei).
- ✓ Sì, perché la loro promozione e affermazioni su questo potenziale DEVONO essere regolamentati.
- ✓ Sì perché bisogna essere consapevoli degli effetti collaterali (tra cui l'automation bias e il deskilling).



GRAZIE!

cabitza @ disco.unimib.it
@cabitzaf

Outcome prediction



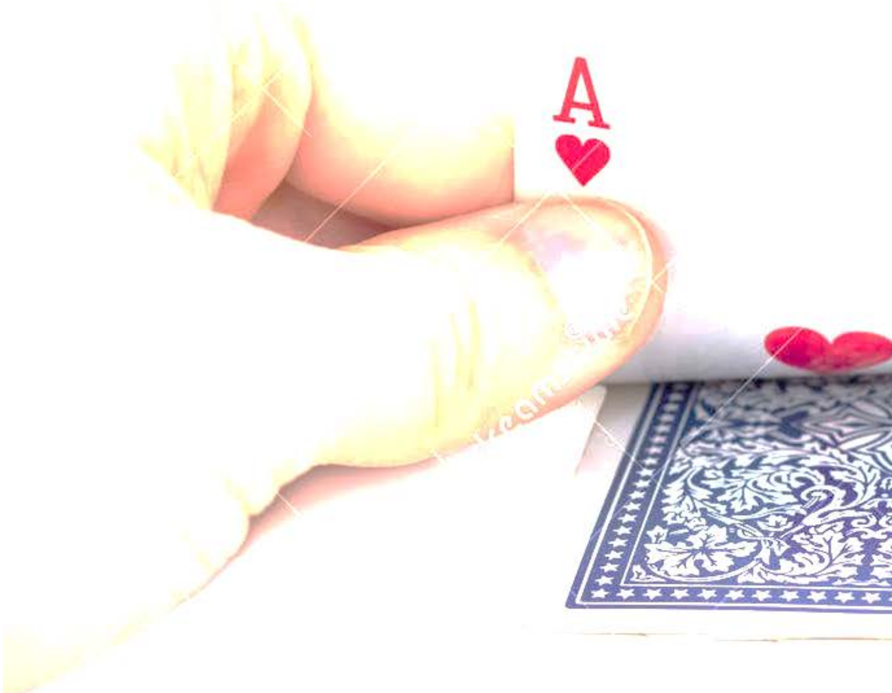
Knee :: Mental score :: 3 months

Hip :: mental score :: 3 months

Knee :: Functional score :: 3 months

Spine :: Pain score :: 3 months

Outcome prediction



Hip:: Mental score :: 3 months

Prediction of the mental score from the SF12 questionnaire (MCS) of hip surgery (first intervention) after 3 months.

Selected features:

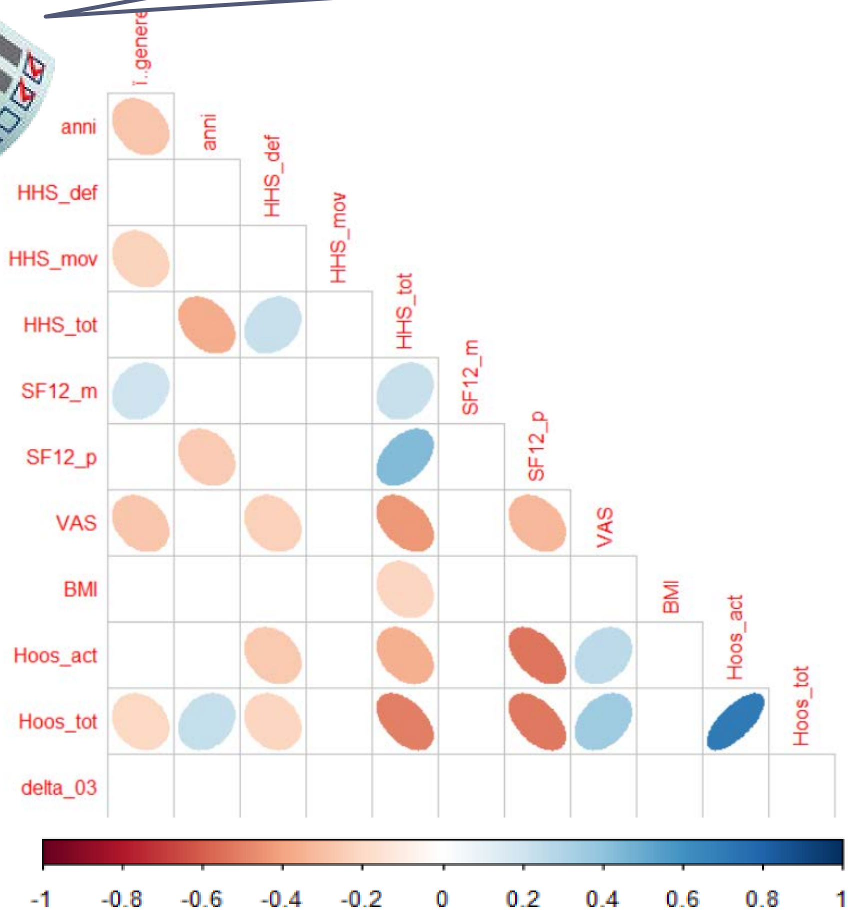
1. Age,
2. gender,
3. pre-op HHS deformity,
4. per-op HSS movement,
5. per-op HSS total
6. pre-op SF12 mental score,
7. pre-op SF12 physical score.
8. Pre-op pain VAS
9. Pre-op BMI
10. Preop Hoos-ps activity

Target variable:

- Delta SF12 mental score (MCS) 0-3 months

Outcome prediction

No cheating: input variables do not correlate with the target variable.



Hip:: Mental score :: 3 months

Prediction of the mental score from the SF12 questionnaire (MCS) of hip surgery (first intervention) after 3 months.

Selected features:

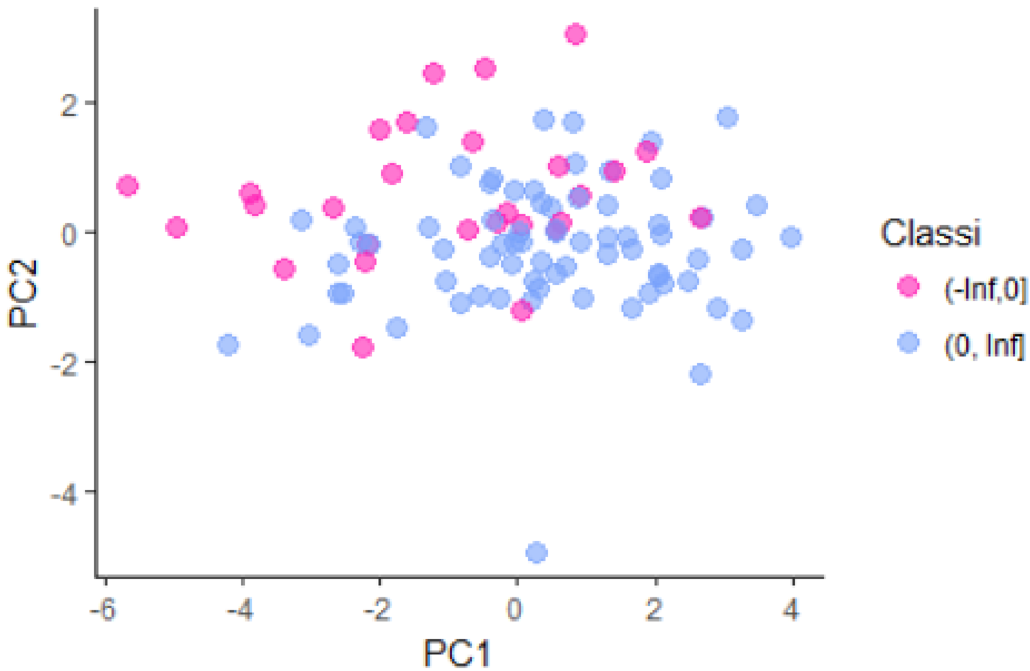
1. Age,
2. gender,
3. pre-op HHS deformity,
4. per-op HSS movement,
5. per-op HSS total
6. pre-op SF12 mental score,
7. pre-op SF12 physical score.
8. Pre-op pain VAS
9. Pre-op BMI
10. Preop Hoos-ps activity

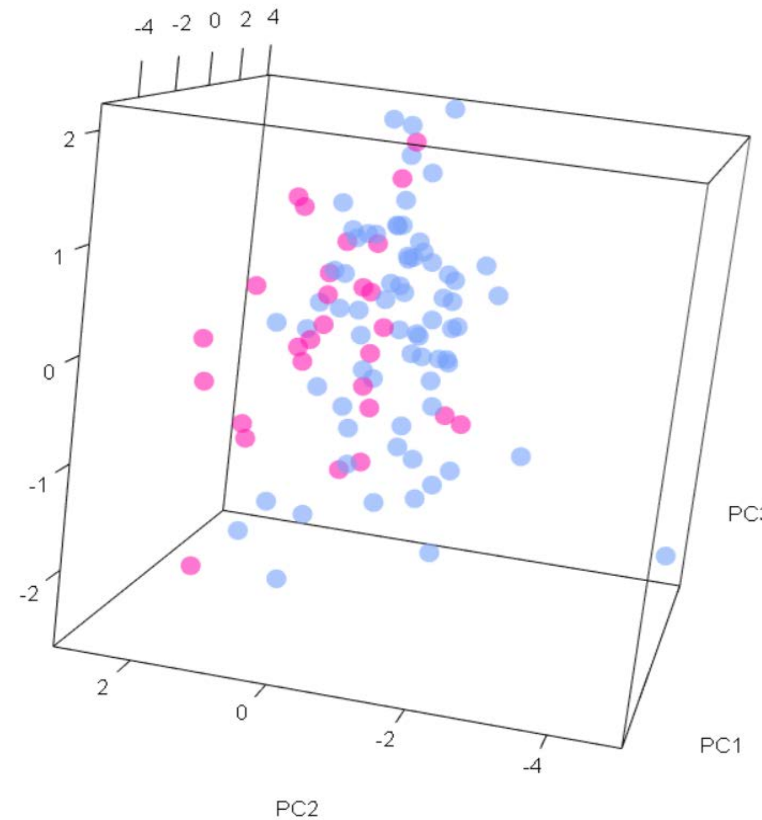
Target variable:

- Delta SF12 mental score (MCS) 0-3 months

OUTCOME PREDICTION

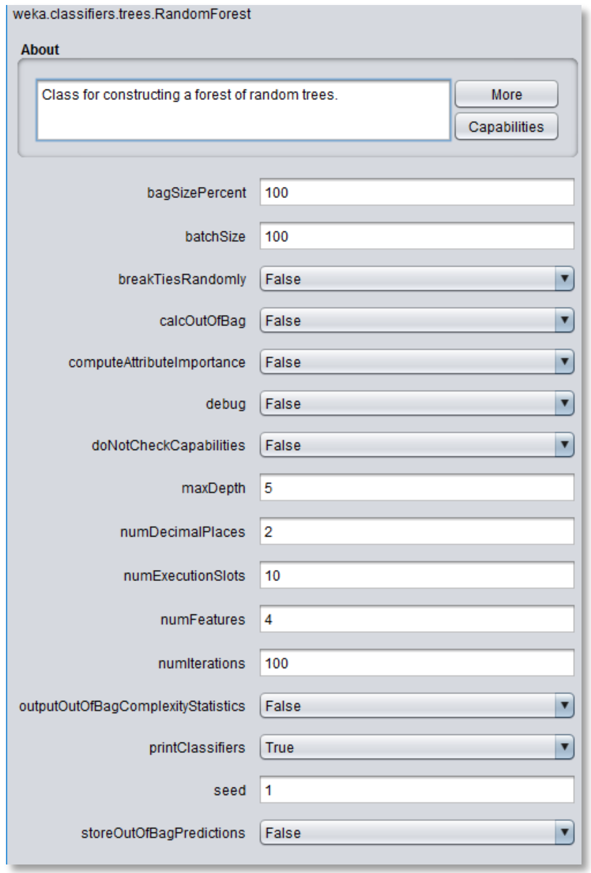
No cheating: worsened patients and those who got better look “mixed”.





Outcome prediction

The most accurate model is the Random Forest.



Hip:: Mental score :: 3 months

Prediction of the mental score from the SF12 questionnaire (MCS) of hip surgery (first intervention) after 3 months.

Selected features:

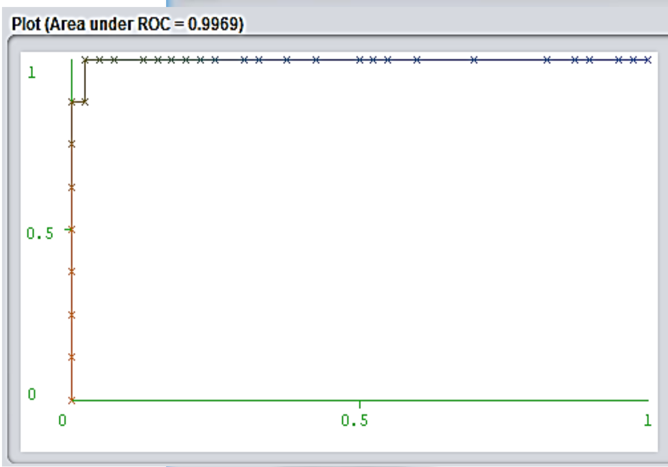
1. Age,
2. gender,
3. pre-op HHS deformity,
4. per-op HSS movement,
5. per-op HSS total
6. pre-op SF12 mental score,
7. pre-op SF12 physical score.
8. Pre-op pain VAS
9. Pre-op BMI
10. Preop Hoos-ps activity

Target variable:

- Delta SF12 mental score (MCS) 0-3 months

Outcome prediction

The most accurate discriminative model is the Random Forest. AUROC: 99%



Accuracy: 97.9%
TP rate: 97.4%
FP rate: 0.04%
Precision: 98.1%
Recall: 97.9%
F-score: 98.0% (on test set)

Hip:: Mental score :: 3 months

Prediction of the mental score from the SF12 questionnaire (MCS) of hip surgery (first intervention) after 3 months.

Selected features:

1. Age,
2. gender,
3. pre-op HHS deformity,
4. -op H
5. p op total
6. pre-op SF12 mental score,
7. pre-op SF12 physical score.
8. Pre-op pain VAS
9. re- BMI
10. reo loos- ctiv

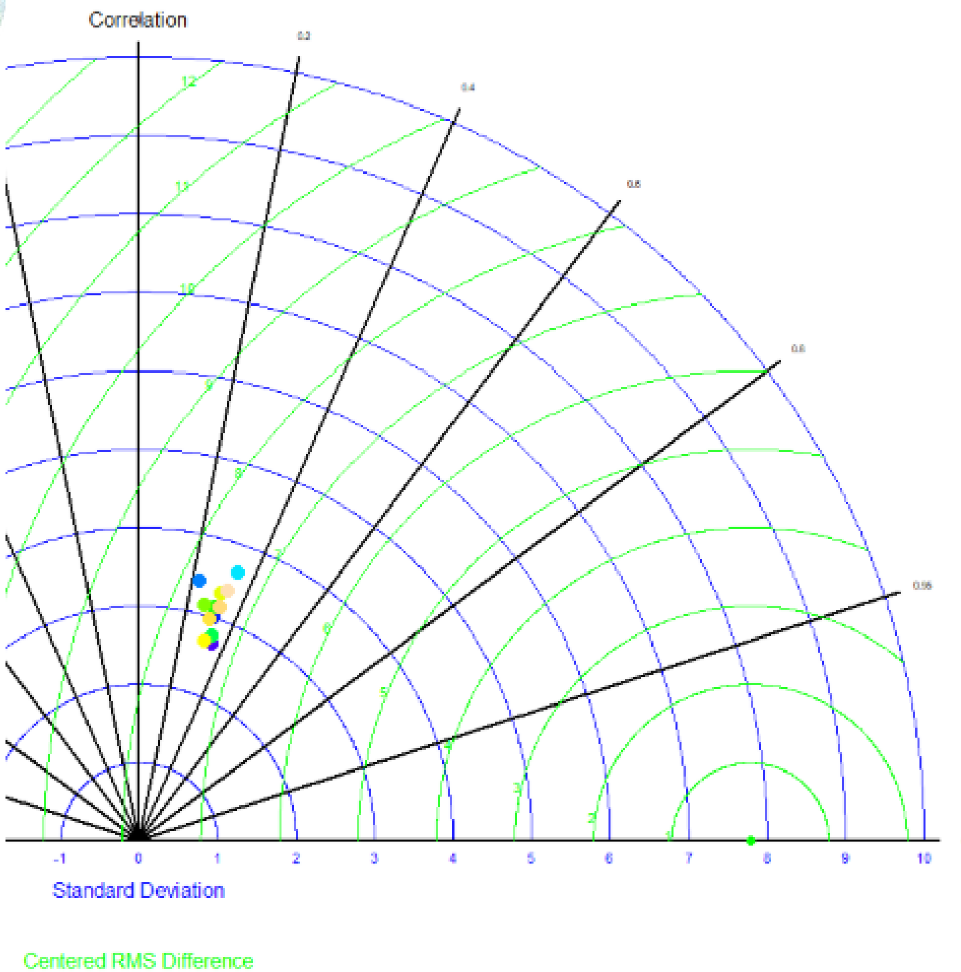
get variable:

➢ Delta SF12 mental score (MCS) 0-3 months

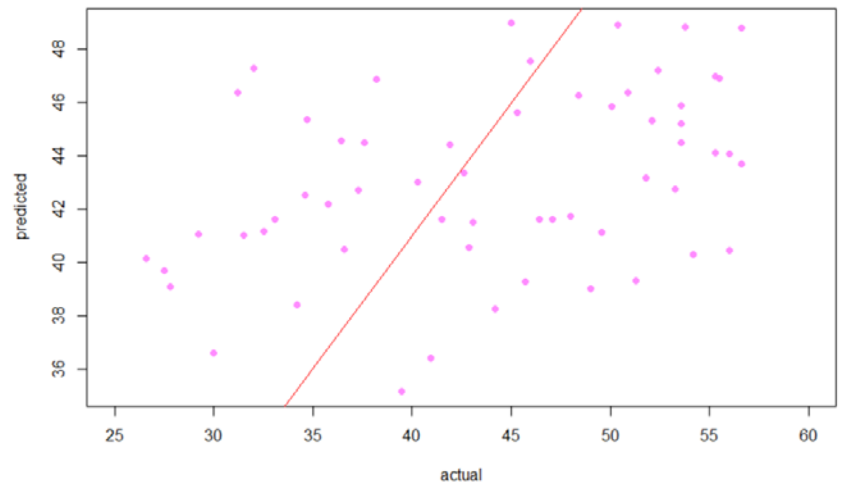
99%

Outcome prediction

The most accurate regression model gives RMSE= 7.5.



Hip:: Mental score :: 3 months



7.5 RMSE