# پیچیدگی های
# یادگیری ماشین و مدلسازی داده محور
## در عصر هوش مصنوعی و یادگیری عمیق

**بابک نجار اعرابی**
دانشکده مهندسی برق و کامپیوتر دانشگاه تهران

دانشگاه علوم پزشکی ایران
شنبه 25 فروردین 1403

# در بستر تحول دیجیتال صحبت می کنیم

- Digital Transformation
  - افزایش تعداد و تنوع سنسورها
  - افزایش قدرت ذخیره سازی اطلاعات
  - افزایش قدرت پردازش اطلاعات
  - توسعه انواع شبکه های ارتباطی

- انبوه سنسور و داده و پردازنده در یک زیست بوم شبکه ای دسترس است

# AI is a
# Disruptive Technology

در همه جا تاثیر خواهد گذاشت

راه فرار ندارد!

Healthcare, Medicine, Business, Education, and ...

# هوش و هوش مصنوعی

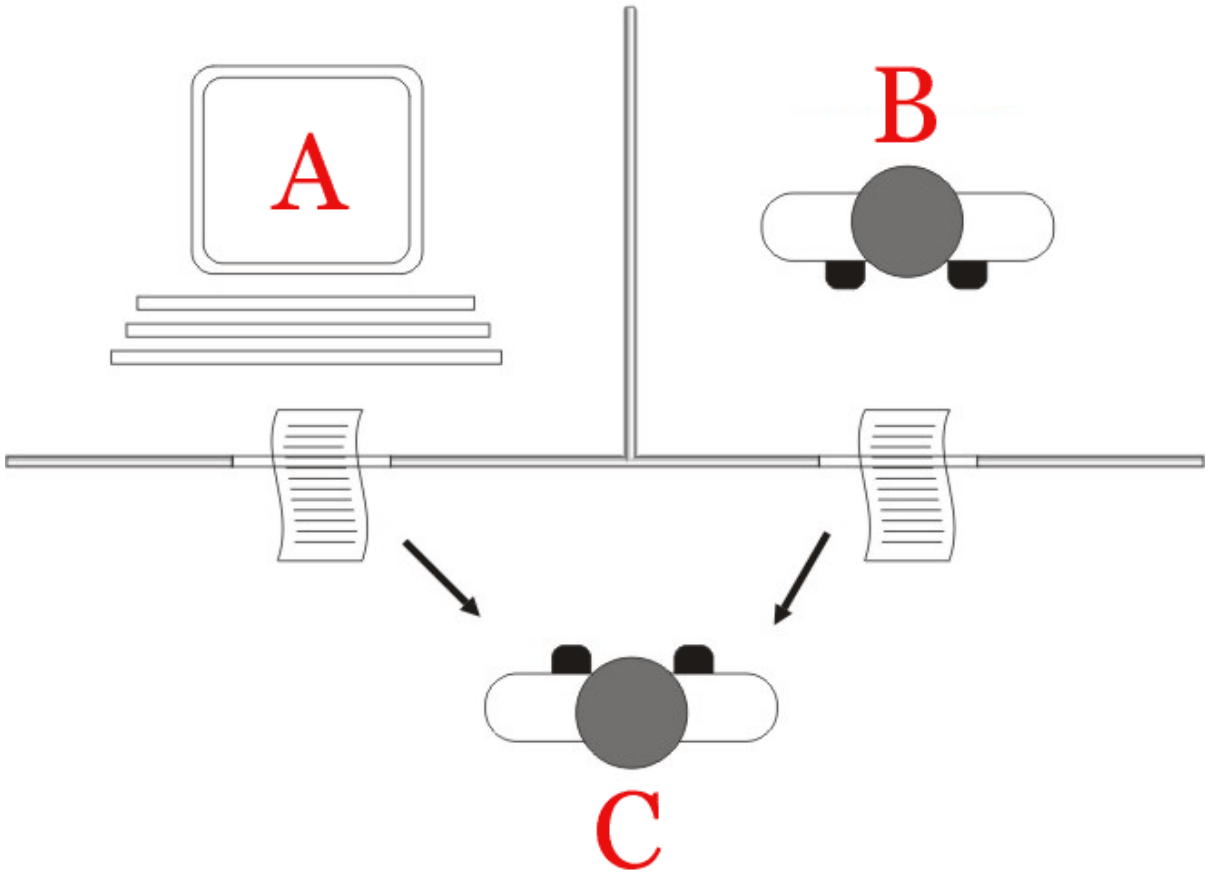
# نگاهی به هوش مصنوعی

- هوش مصنوعی سمبولیک ← GOFAI

- 1950-1985
- High level Rep. of Problems
- High level programing languages: LISP, Prolog
- Logic
- Search
- IBM Deep Blue Chess Computer (1997)

- هوش چیست؟
- چه ماشینی هوشمند است؟
- تعبیر حداقلی
- تعبیر آلن تورینگ

# Turing Test



# یادگیری ماشین در بستر هوش مصنوعی

**… What we want is a machine that can learn from experience**

Alan Turing 1947

# یادگیری و یادگیری ماشینی

# یادگیری در کودک


- درک سیستم موتوری
- کنترل حرکت
- درک سیستم صوتی
- یادگیری زبان

- همه از راه تعامل با محیط، مشاهده، آزمایش، آزمون و خطا، اکتشاف فضای ادراکی و فضای اعمال

# ماشین با توان یادگیری از آزمایش ها

- مشاهده
- اندازه گیری
- داده ها

- **Machine that Learns from Data**
  - وقتی نمی توان یک الگوریتم سر راست برای حل مساله یافت
  - روش های آماری (Statistics)
  - شبکه های عصبی مصنوعی

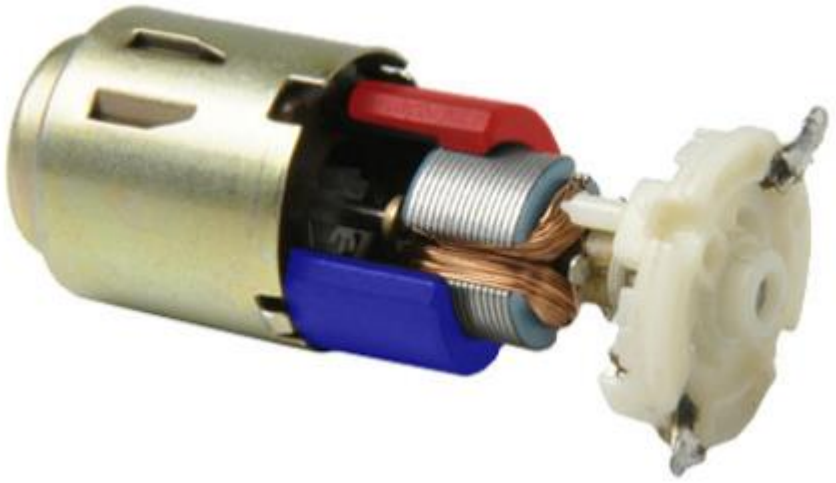
# Modelling

# اهمیت داشتن مدل

- درک و تحلیل پدیده ها بدون دخل و تصرف در آنها
- پیش بینی: حرکت به جلو در زمان
- فیلتر کردن: حرکت به عقب و یافتن ریشه های وضع موجود
- شبیه سازی: درک سازوکارهای درونی پدیده ها
- کنترل: مداخله برای رسیدن به وضع مطلوب
- عیب یابی: مدلسازی وضعیت سالم و معیوب و تمیز آنها
- ...

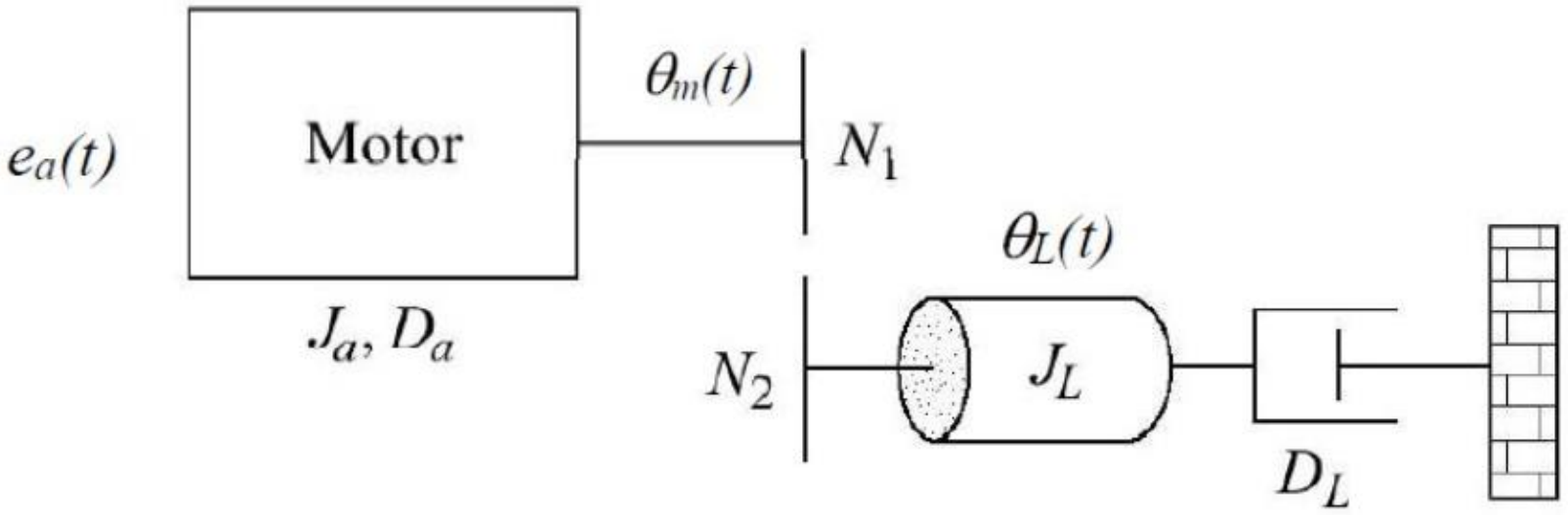
# White-box Modelling

# DC Motor



Electrical parts
Mechanical parts
Electro-mechanical coupling

# DC motor, load and torque-speed




$$J_m = J_a + \left(\frac{N_1}{N_2}\right)^2 J_L; \quad D_m = D_a + \left(\frac{N_1}{N_2}\right)^2 D_L; \quad \theta_m(t) = \left(\frac{N_2}{N_1}\right)\theta_L(t)$$

$$e_a(t) = R_a i_a(t) + K_b \frac{d\theta_m(t)}{dt}; \quad T_m(t) = K_a i_a(t)$$

where $e_a(t)$ and $i_a(t)$ are the armature voltage and current, $J_m$ and $D_m$ are the equivalent inertia and damping; $T_m(t)$ is the torque developed by the motor; $N_1/N_2$ is the ratio of the gear system.)

# Closer look at equations

***Electrical part*:**

$$e_a(t) = R_a i_a(t) + K_b \frac{d\theta_m(t)}{d(t)}$$

***Mechanical load*:**

$$T_m(t) - D_a + \left(\frac{N1}{N2}\right)^2 D_L \frac{d\theta_m(t)}{dt} = J_a + \left(\frac{N1}{N2}\right)^2 J_L \frac{d^2\theta_m(t)}{dt^2}$$

***The electrical − mechatronic part*:**

$$T_m = K_a i_a(t)$$

# White-box Analytical Model

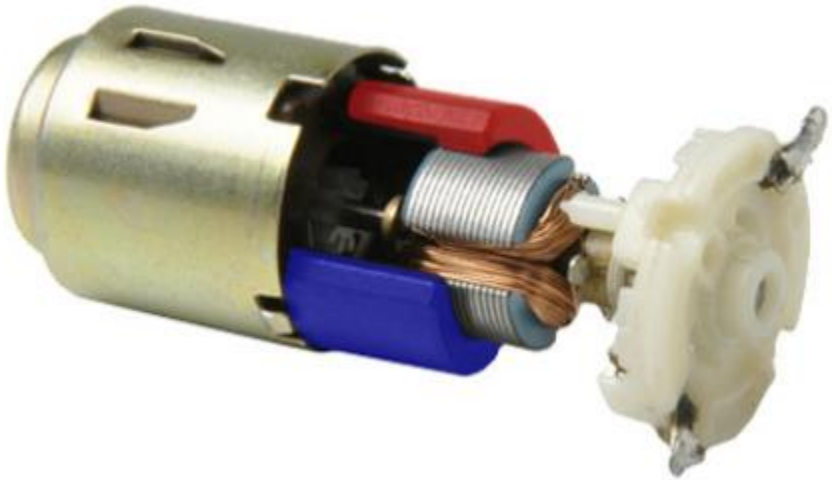
- **one can clearly explain**
  - **how they behave**
  - **how they produce outcomes**
  - **what are the influencing variables**
- **A white-box model is explainable by design**
- **Keyword: Explainable**

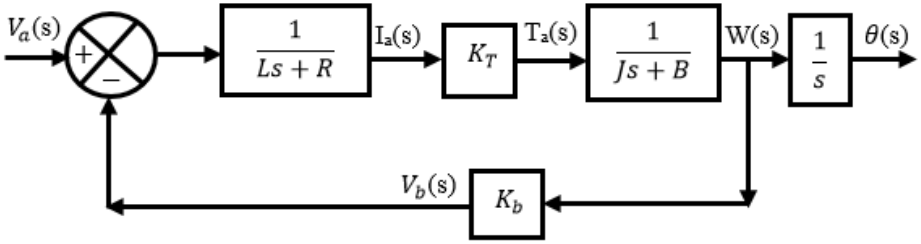
# Black-box Modelling

# DC Motor

Electrical parts
Mechanical parts
Electro-mechanical coupling

$$G(s) = \frac{k}{s(s^2 + as + b)}$$





# Black-box Modelling of DC Motor

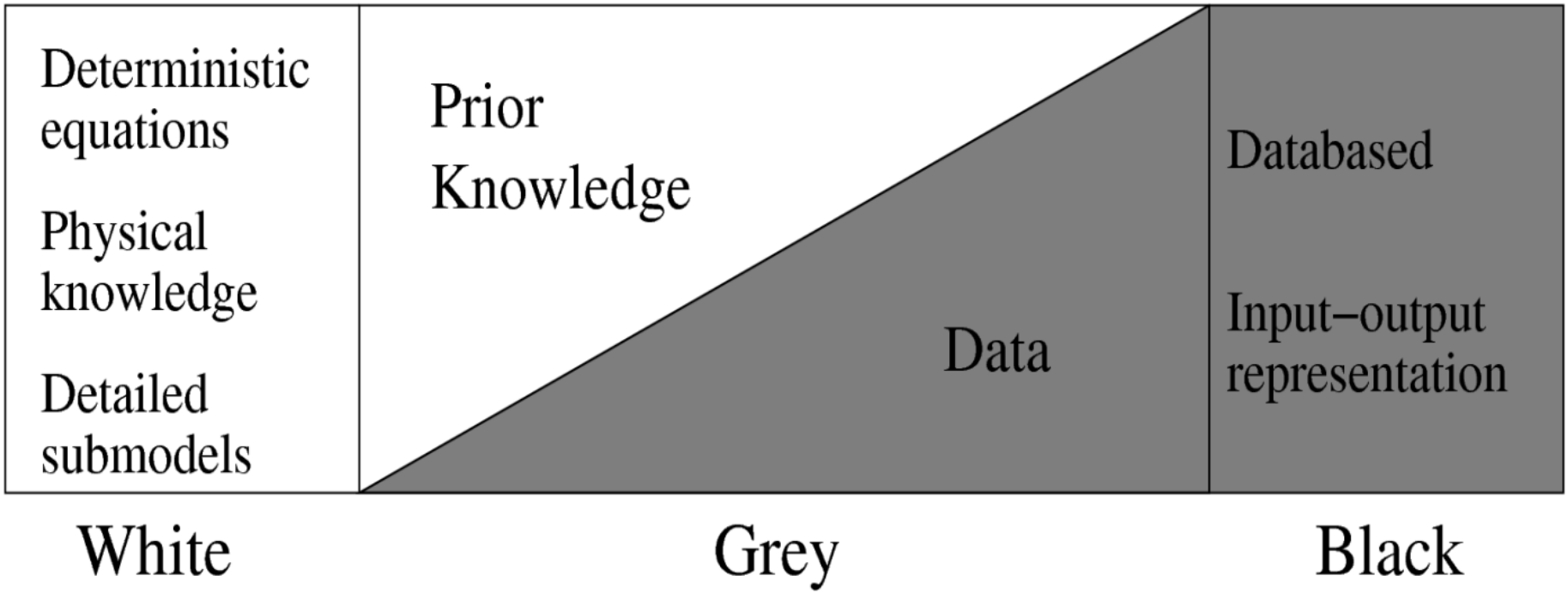
- Design an experiment to estimate Three parameters: $a$, $b$, $k$

$$G(s) = \frac{k}{s(s^2 + as + b)}$$

- What we learn in system identification
- Presumably: non need to know anything about inside the system
- Zadeh's definition of system identification

# Grey-box Modelling

# Prior Knowledge + Data



# White-box Data-driven Modelling

# White-box Data-driven Model

- **A white-box model is explainable by design**
- **One can clearly explain**
  - **How they behave**
  - **How they produce outcomes**
  - **What are the influencing variables**

- **What make a data driven model close to white-box?**
  - **Features/Regressors have to be Understandable**
  - **The ML process has to be Transparent**

# Then Why Not Stick to While-box Modelling?

# Let stick to White-box modelling

- It is not always possible!
  When we deal with Large scale/complex system
  - Lots of inputs and/or outputs
  - Complex mapping or complex internal dynamics
  - Lots of internal variables
  - Different forms of uncertainty
    Noise, disturbance
    Unknown/unmodelled dynamics/subsystems
    unknown input variables

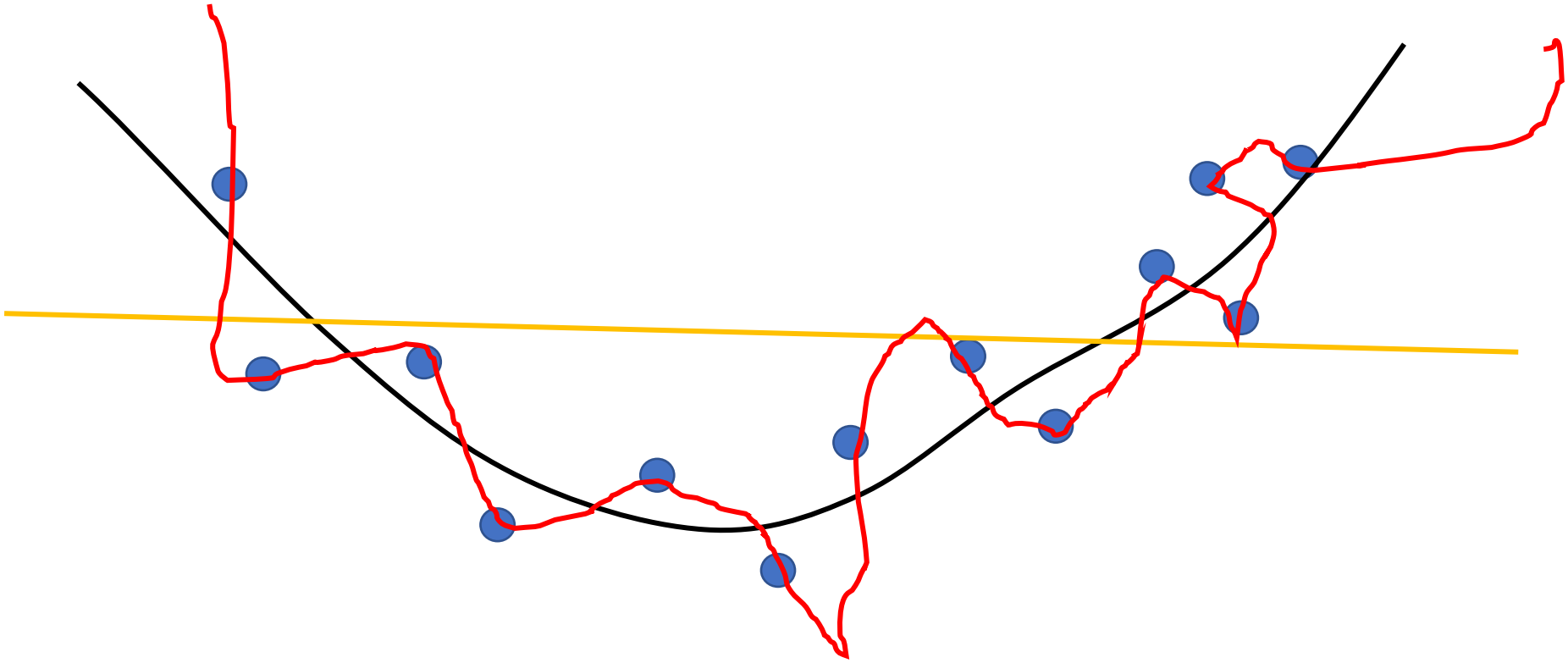
# Then Why Not Stick to Black-box Modelling?

# توقع حداکثری

- اگر سیستم را خوب و به مدت کافی در حالات مختلف مشاهده و ورودی ها و خروجی های آن را اندازه گیری کنیم، تمام اطلاعات درباره سیستم در دسترسمان است.
- در سال های اخیر تمایل به روش های محاسباتی گسترش یافته است
  - افزایش تعداد و تنوع سنسورها
  - افزایش قدرت ذخیره سازی اطلاعات
  - افزایش قدرت پردازش اطلاعات
  - توسعه انواع شبکه های ارتباطی
- انبوه سنسور و داده و پردازنده در دسترس است

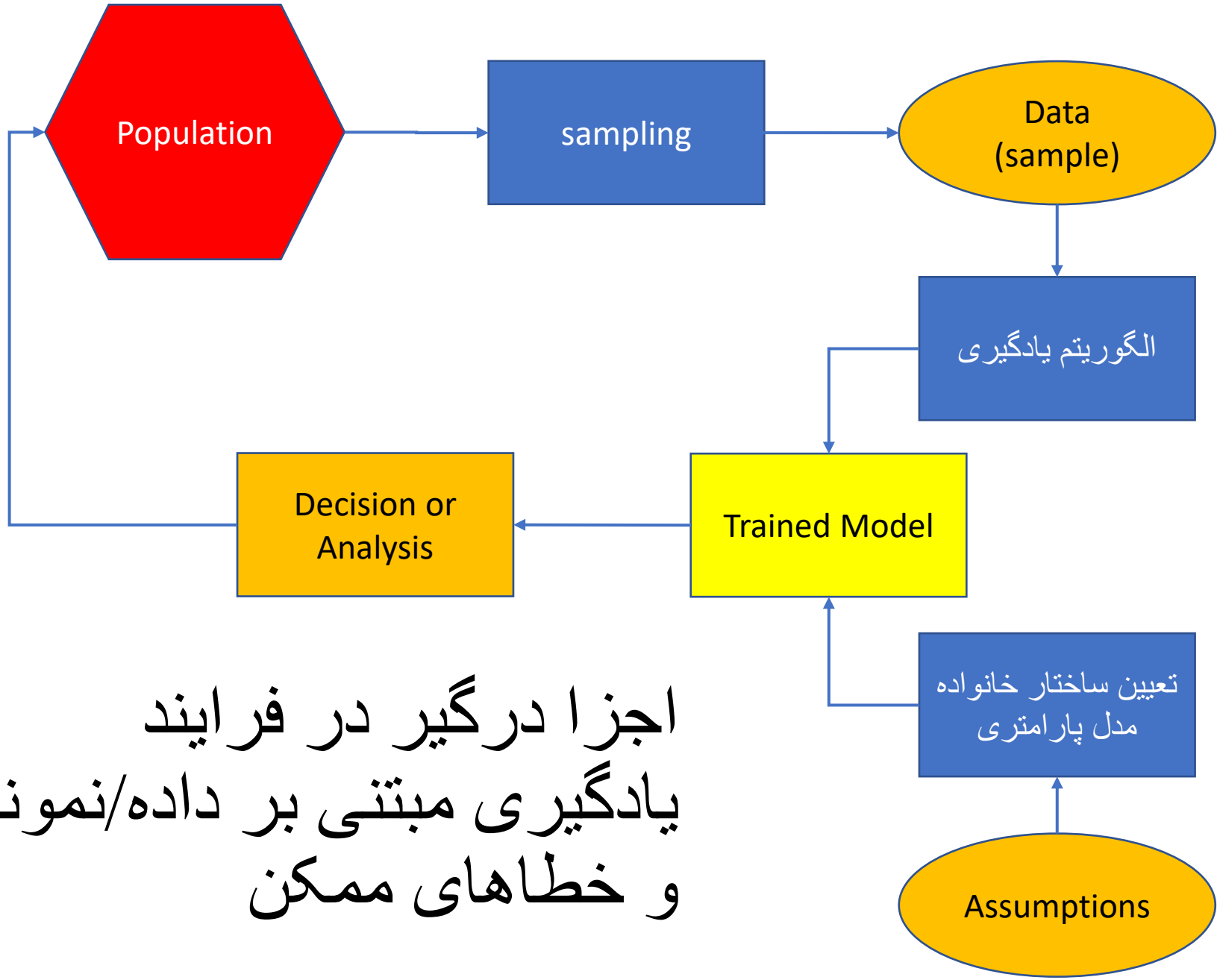
# چندان هم ساده نیست

- مثال از برازش منحنی به داده ها



# سوال اساسی از منظر آماری

- مایلیم گزاره هایی بیان کنیم که درباره یک جمعیت حاوی اطلاعات دقیق و درستی باشد
- اما تنها بخشی از جمعیت را مشاهده کرده ایم: نمونه (sample)
- چه کنیم تا تصمیم ها/تحلیل های مبتنی بر sample قابل تعمیم به جمعیت باشد؟
- بحث استنباط آماری




اجزا درگیر در فرایند
یادگیری مبتنی بر داده/نمونه
و خطاهای ممکن

# Let stick to Black-box modelling

- How about Generalization?
  - Don't we memorize the data?
  - Do we get a good representation?
  - Do the model tolerate variations? (Invariance)
  - Do the model tolerate shift of distribution? (Transfer learning & Domain adaptation)
- How about Explainability?
  - Understandable features/regressors
  - Transparent ML process
- How about Fairness?
  - How to cope with database bias?

# آیا ماشین می تواند تصمیم های خوبی بگیرد؟

بحث بایاس ها

# Database Bias

I hate bias

small sample size
unfair sample

# Inductive Bias

I love bias

که عشق آسان نمود اول ولی افتاد مشکل ها

# Bias-variance Trade-off

**Model complexity**

**Generalization**

# Nonstationary condition

**Gradual Learning**

**Incremental Learning**

**Sequential Learning**

**Adaptive Model**

**Evolving Model**

# آیا ماشین می تواند بهتر از انسان تصمیم بگیرد؟

شغل هایی که از بین می رود/ایجاد می شود/تغییر می کند

# Prospects of AI in Medicine (Radical views, 2012)

## Vinod Khosla: Machines will replace 80 percent of doctors (by 2035)

By **LIAT CLARK**

*Tuesday 4 September 2012*

**Machines will replace 80 percent of doctors in a healthcare** future that will be driven by entrepreneurs, not medical professionals, according to Sun Microsystems co-founder Vinod Khosla.

Khosla, who wrote an article entitled *Do We Need Doctors Or Algorithms?* earlier this year, made the controversial remarks at the Health Innovation Summit in San Francisco, hosted by seed accelerator Rock Health. The article had already touched on some of the points of his keynote speech, however it was at the summit

https://www.wired.co.uk/article/doctors-replaced-with-machines#:~:text=Machine%20learning%2C%20he%20argues%2C%20will,percentile%20can%20remain%2C%20apparently)+

# Prospects of AI in Medicine (Radical views, 2016)

- "I think that if you work as a radiologist you are like coyote that has already run over the edge of the cliff, but hasn't yet looked down so doesn't realize there's no ground underneath him. People should stop training radiologists right now. It's just completely obvious that within five years deep learning is going to do better than radiologists. It might be ten years."

Geoffrey Hinton

# Prospects of AI in Medicine (Radical views, 2017)

**Andrew Ng** 
@AndrewYNg

Should radiologists be worried about their jobs? Breaking news: We can now diagnose pneumonia from chest X-rays better than radiologists.
stanfordmlgroup.github.io/projects/chexn...

3:20 PM · Nov 15, 2017 from Mountain View, CA · Twitter Web Client

**1,344** Retweets **162** Quote Tweets **2,274** Likes

# آیا بترسیم؟

موضوع آنقدر ها هم ساده نیست.

متفکر و هوشیار باشیم.

Augmented AI

Doctor is the Boss!

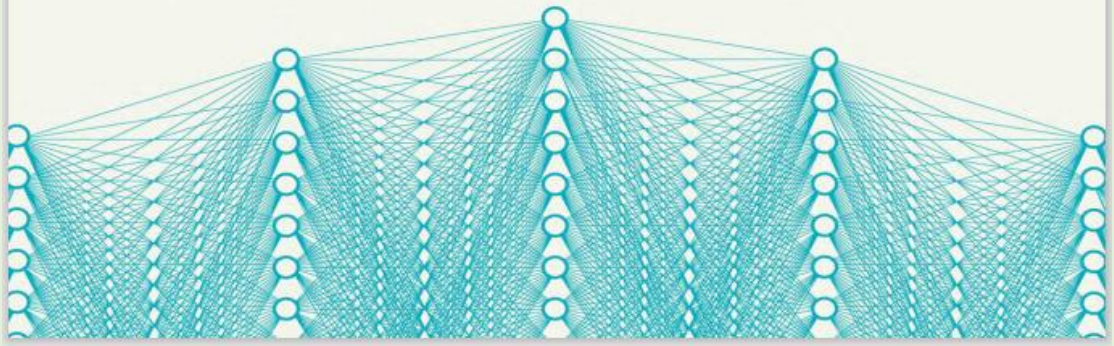
# Let Introducing a Good Book

# THE ALIGNMENT PROBLEM

Machine Learning and Human Values

BRIAN CHRISTIAN

Best-Selling Author, *Algorithms to Live By*



# Supervised Learning
# Reinforcement Learning
# Inverse RL

**The alignment problem**

# Are decision made by ML based AI algorithms aligns with human values?

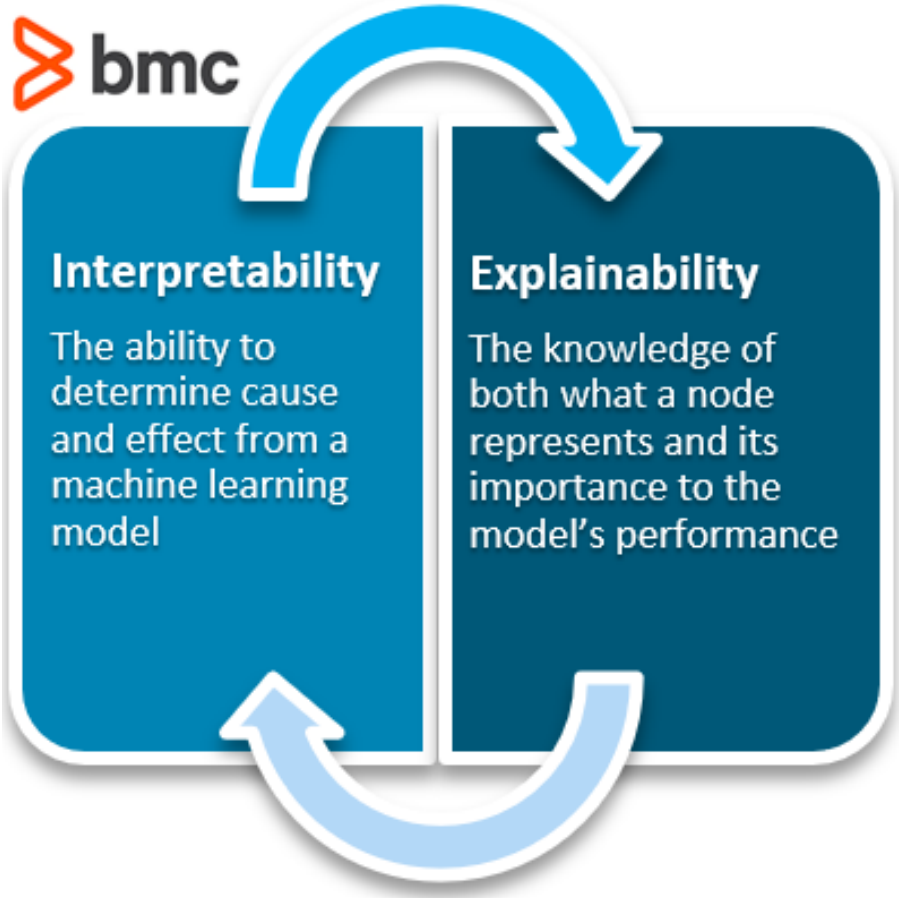
- Prophecy: ANN approach in ML
  - COMPAS: Correctional Offender Management Profiling for Alternative Sanctions. Case management and decision support tool used by U.S. courts to assess the likelihood of a defendant becoming a recidivist. Bias towards certain demography. Lack of transparency.
- Agency: RL and the psychological study of reward
  - Importance of curiosity, in which RL agent intrinsically motivated to explore their environment, rather than exclusively seeking the external reward
- Normativity: Imitative Learning and Inverse RL (IRL)

# COMPAS case

- Blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend, whereas COMPAS makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower-risk but go on to commit other crimes
- COMPAS software is somewhat more accurate than individuals with little or no criminal justice expertise, yet less accurate than groups of such individuals. On average, they got the right answer 63 percent of their time, and the group's accuracy rose to 67 percent if their answers were pooled. COMPAS, by contrast, has an accuracy of 65 percent.

# Interpretability
# &
# Explainability

# Interpretability & Explainability




Sometimes they used interchangeably in the literature

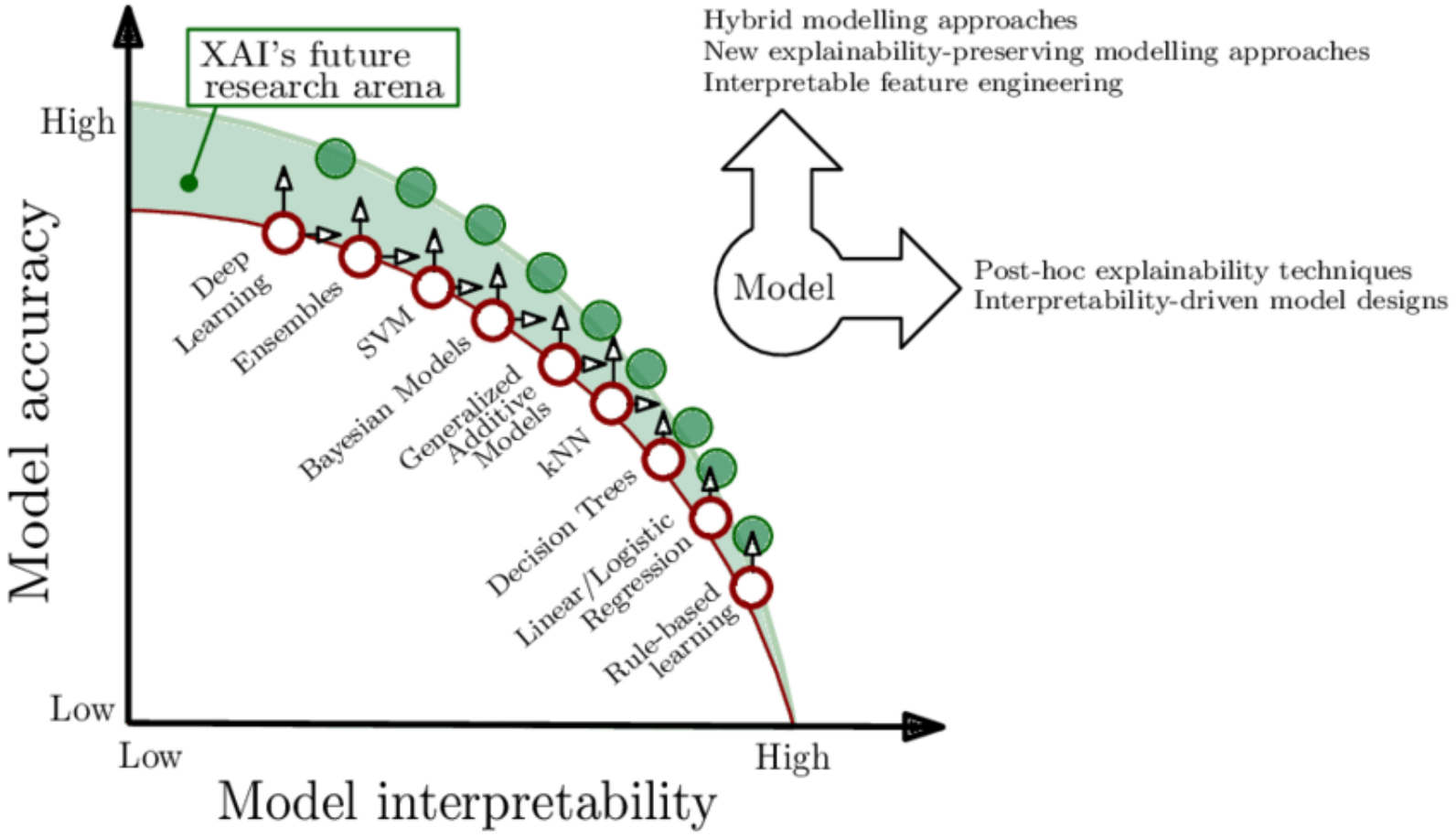
# Terminology

- Nowadays, Explainable Artificial Intelligence (xAI) is used as a umbrella term for collective concerns and challenges related to
- Explainability
- Interpretability
- Fairness
- …

# Current Trend in AI:

# Explainable Artificial Intelligence (xAI)

# Interpretability vs Accuracy: An xAI Trade-off

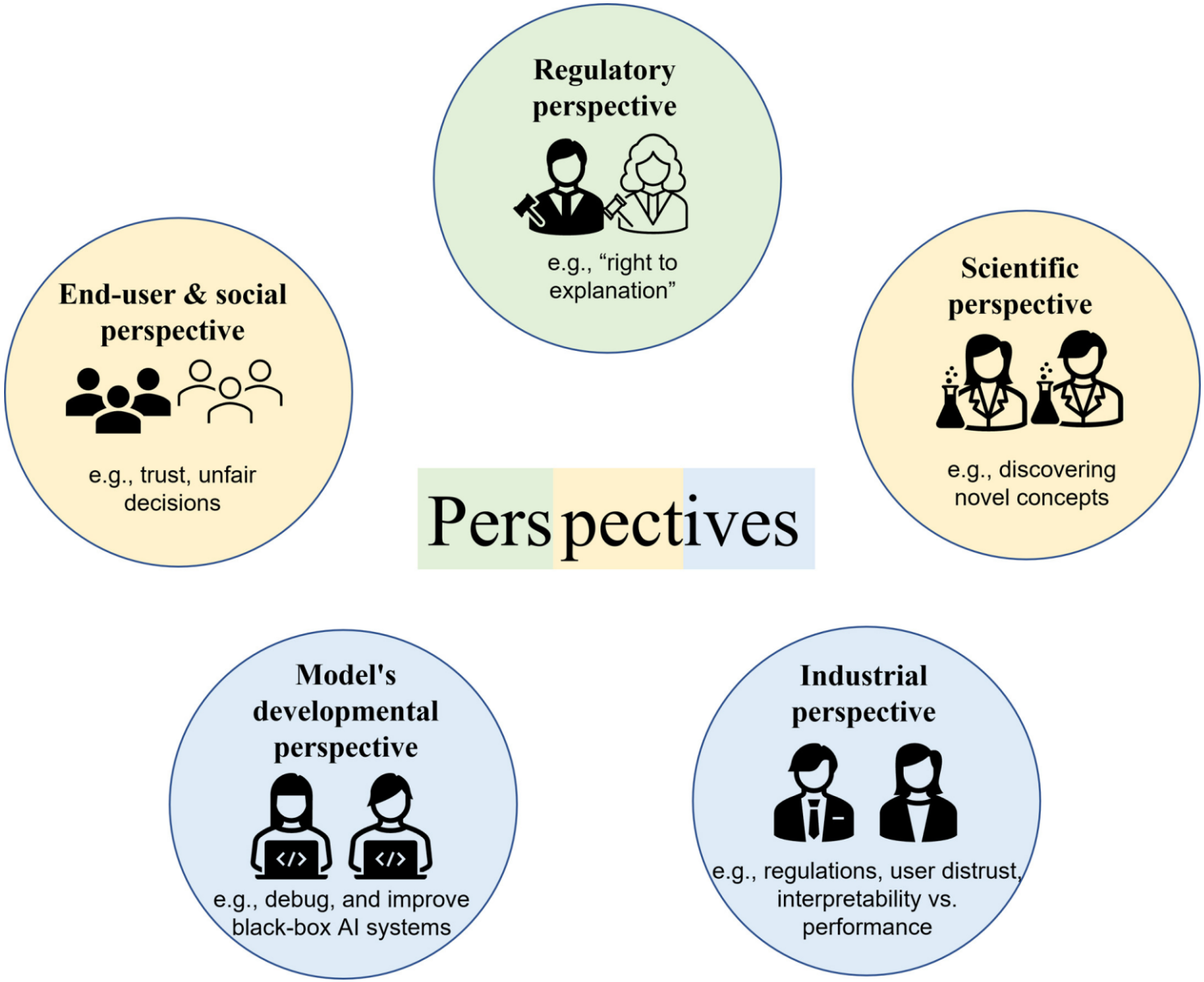


# Benefits of Explainable AI (xAI)

- **Reducing Cost of Mistakes**
  - Decision-sensitive fields: Medicine, Finance, Legal, etc.
  - Allow for expert supervision
- **Reducing Impact of Model biasing**
  - Gender Bias for Apple Cards
    Women received less credit than their spouses even though they shared the same income and credit score
  - Racial Bias by Autonomous Vehicles (AI Self-driving cars) how effective various "machine vision" systems are at recognizing pedestrians with different skin tones
  - Gender and Racial bias by Amazon Rekognition --> computer vision platform (both Pre-trained algorithms and user custom dataset) Rekognition performed worse when identifying an individual's gender if they were female or darker-skinned

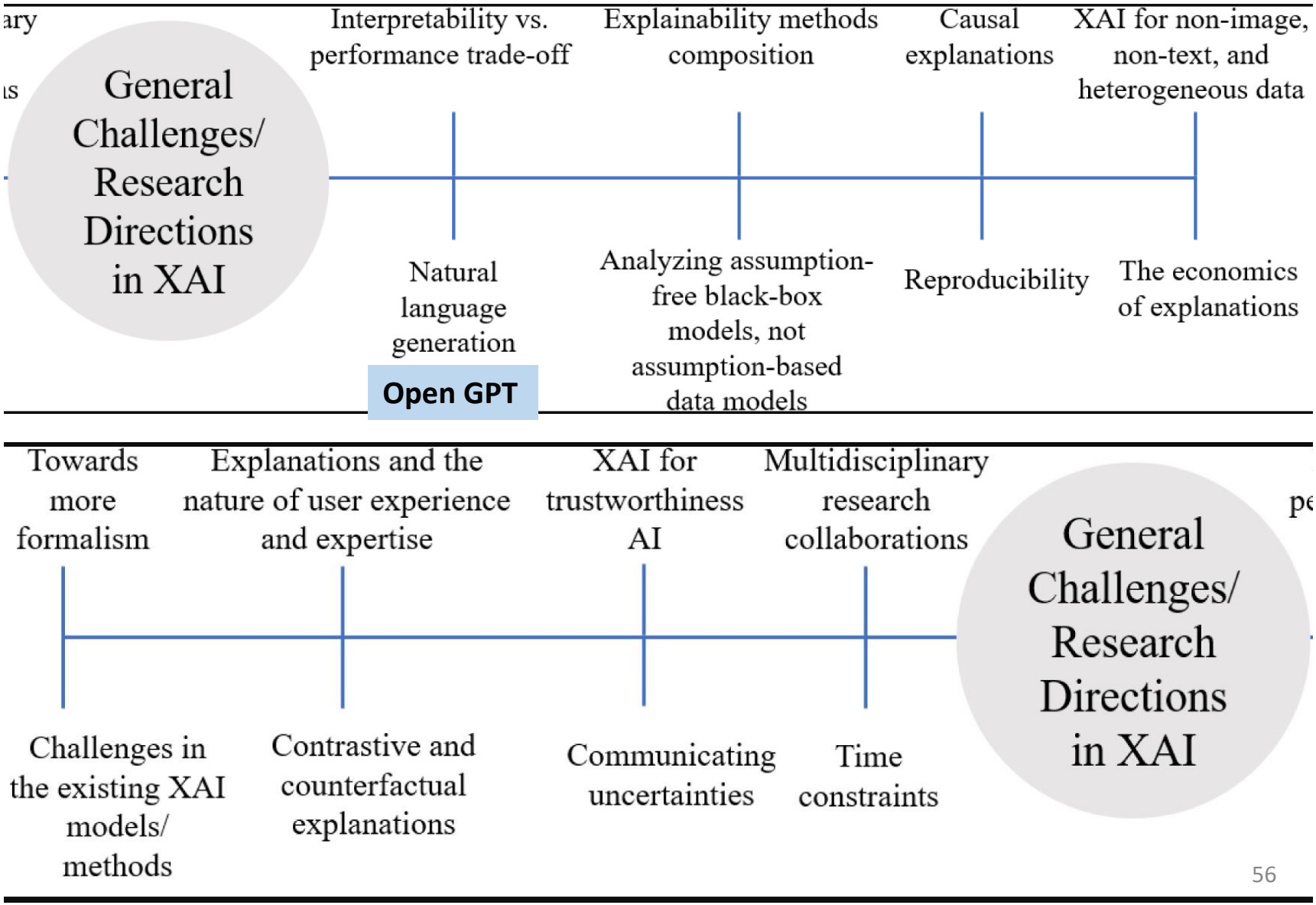
# Benefits of Explainable AI (xAI)

- **Responsibility and Accountability**
  - enabling a person who can be responsible and accountable for decisions and errors (e.g. auto. driving)
- **Code Confidence**
  - Every inference, along with its explanation, tends to increase the system's confidence
- **Code Compliance**
  - to comply with the regulatory bodies' pressure for more explainable AI-based products




Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities, Knowledge-Based Systems, Volume 263, 5 March 2023, 110273

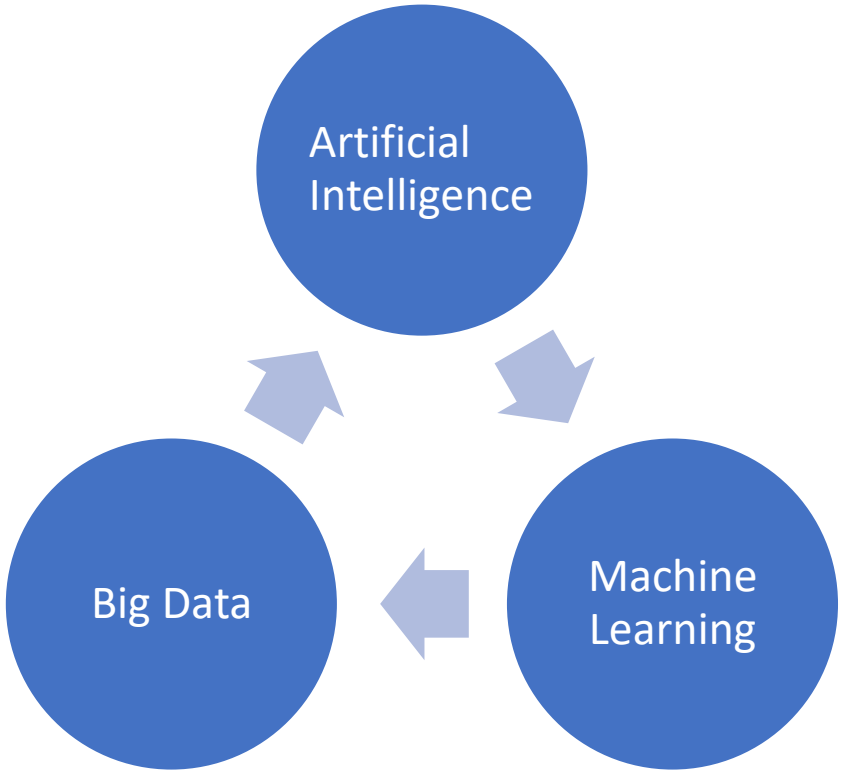
# General Challenges and Research Directions in xAI



# Effect of Money in ML Research

# Big players

- Big companies make lots of money in that golden triangle
- Microsoft
- Alphabet
- Amazon
- IBM
- Alibaba
- Baidu
- …



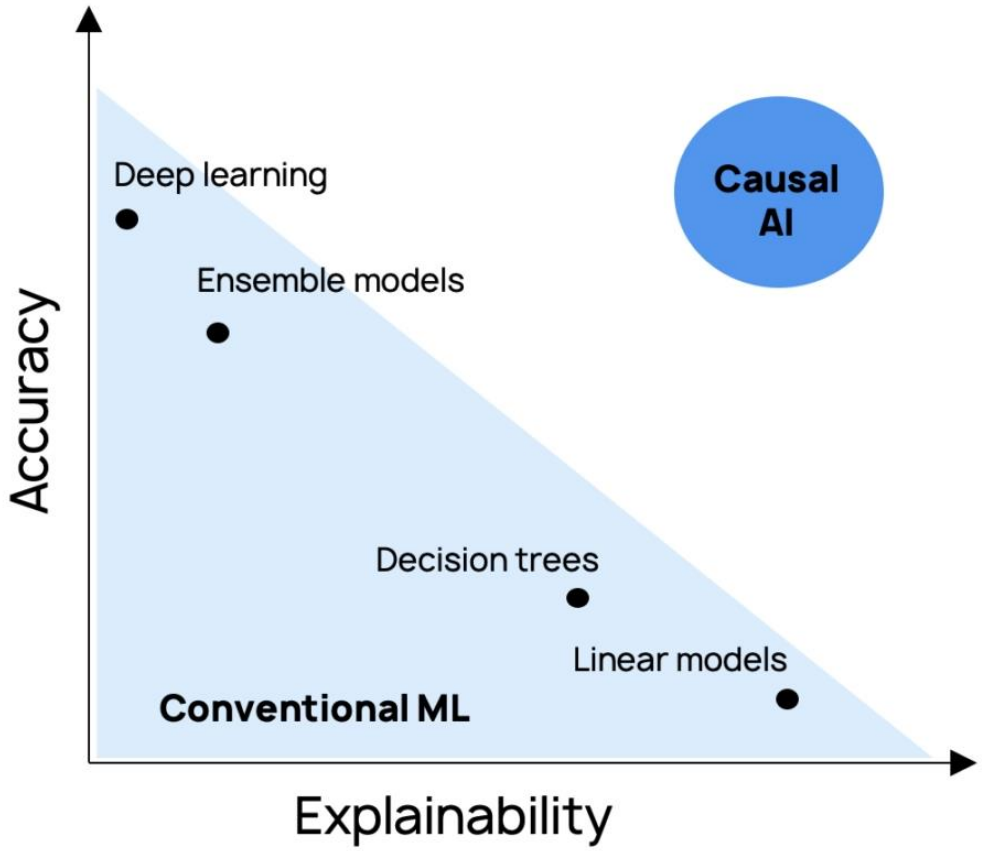
# They like tools that perform

- Tendency towards performance
  In expense of
- Explainability
- Interpretability
- Fairness

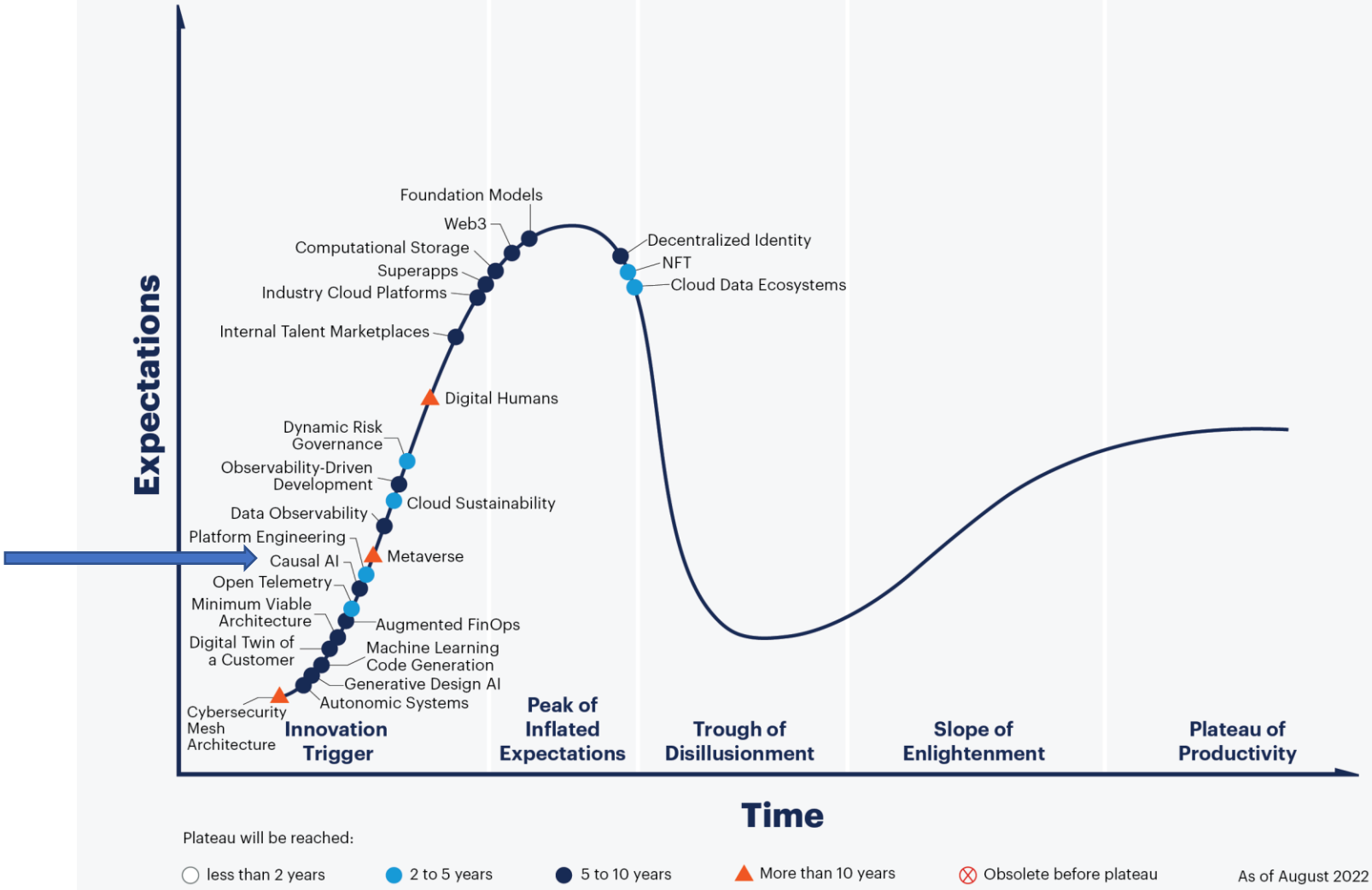
- An important ethical challenge for scientists

# Causal AI
## Causal Models
## &
## Causal Representation Learning

# xAI does not explain enough!



# Hype Cycle for Emerging Tech, 2022




gartner.com

Source: Gartner



Gartner

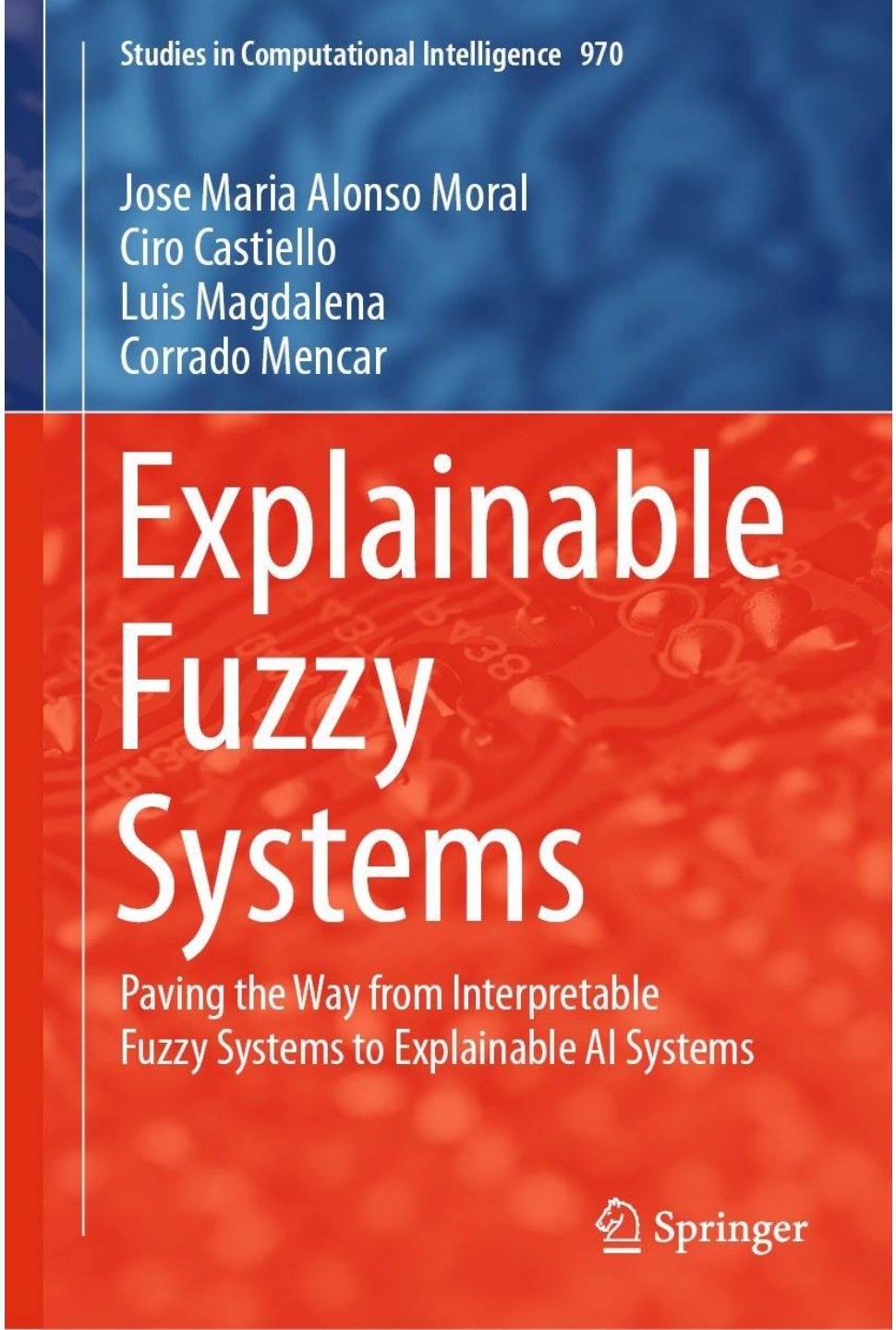
# Fuzzy Models
# &
# Approximate Reasoning

# Rule-based fuzzy systems

- Explainable
- Interpretable
- Clear Decision making process
- Allow for human expert supervision
- Allow for hybrid data + expert knowledge

- How about heuristics?

AI community put an emphasis on this topic



# Representation

# xAI and Representation

- A good representation means
  - Simple decision-making machine
  - High confidence in decision
  - Explainable decision
  - Accurate decision (as much as possible)

- Representation Learning
  - along with (or instead!)
    decision making learning (Classification & Regression)
- Metric Learning
  - Deep metric Learning
  - Similarity Learning

# Parsimony

# Let do not forget the basics!

- whenever we have different explanations of the observed data, the simplest one is preferable (Occam's razor)
- Noone is after complexity in machine learning
- Everybody is after better generalization

- Parsimonious Neural Network (PNN): combine neural networks with evolutionary optimization to balance accuracy with parsimony, Nature's Scientific Reports, 2021

# کلام آخر

- دقت بدست آمده به کمک یادگیری با سرپرستی چشمگیر است
- مدل های مولد مبتنی بر مبدل ها مسایلی را حل می کنند قبلا حل پذیر نمی نمود (مثل Chat-GPT)
- توضیح پذیری، انصاف در تصمیم، بایاس خوب و بد، تعمیم پذیری، علیت، جنبه های حقوقی و اخلاقی سپردن تصمیم به ماشین ها سوالات باز مهم این حوزه هستند
- پزشکان معزز درباره آینده حرفه ای خود تحت تاثیر هوش مصنوعی، نگران خیر اما هوشیار باشند. قطعا نیاز به باز آموزی و توسعه دانش هست

# ممنون از توجه شما

بابک نجار اعرابی

دانشکده مهندسی برق و کامپیوتر

دانشگاه تهران

araabi@ut.ac.ir