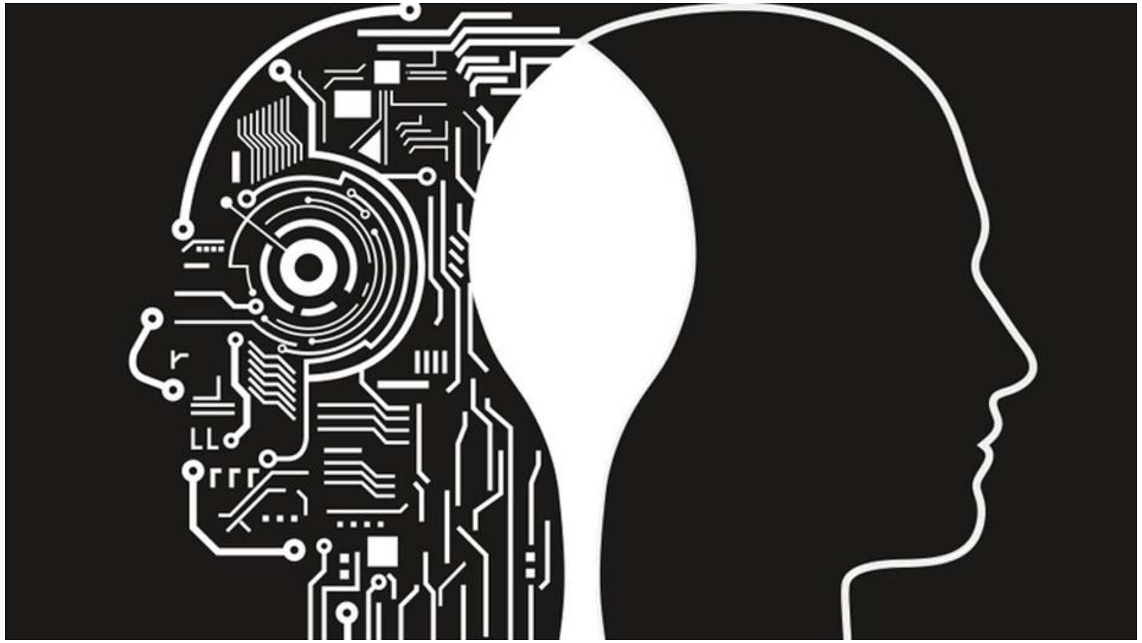
Artificial Intelligence & Artificial Ignorance

Georg Gottlob

University of Oxford & Unical (soon)



https://businessjournaldaily.com/artificial-intelligence-for-good-and-for-bad/

Funded by: The RAISON DATA Project – The Royal Society

The Development of AI from its Beginnings

Two mostly independent and competing approaches:

SYMBOLIC AI
(LOGIC, KR & REASONING)

For 1000s of years: Formal Logic: Aristotle, Avicenna, Llull, Duns Scotus, Occam, Leibniz, Hilbert, Frege, Russel&W., Gödel, Tarski…

1956: Birth of AI - John McCarthy: **KR via Logic**; Newell, Simon, Shaw: Theorist

1964: **Expert systems** - Ed Feigenbaum: Dendral & Mycin(1972); Lenat: CYC (KR)

Currently: **Expressive and efficient logics to work with Big data**. Ontological reasoning (Description logics, Datalog variants), Probabilistic Reasoning Markov Logic, etc.

SUBSYMBOLIC AI
(NEURAL NETWORKS)

1943 McCulloch & Pitts modelled simple **neural networks** by electrical circuits

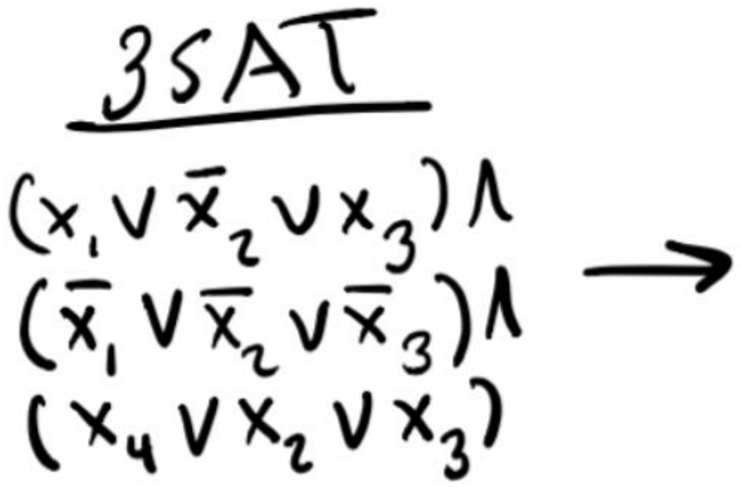
Frank Rosenblatt: **Multilayer Perceptron**

1986 Rumelhart, Hinton & Williams **Backpropagation**, basis for efficient ML

Currently: **Great success of NN-based ML**: New architectures, applications in game playing (AlphaGo), pattern recognition, ChatGPT etc, but also shortcomings…

More Recent Progress in <u>Symbolic</u> AI

Constraint Processing & SAT solving (NP-hard)



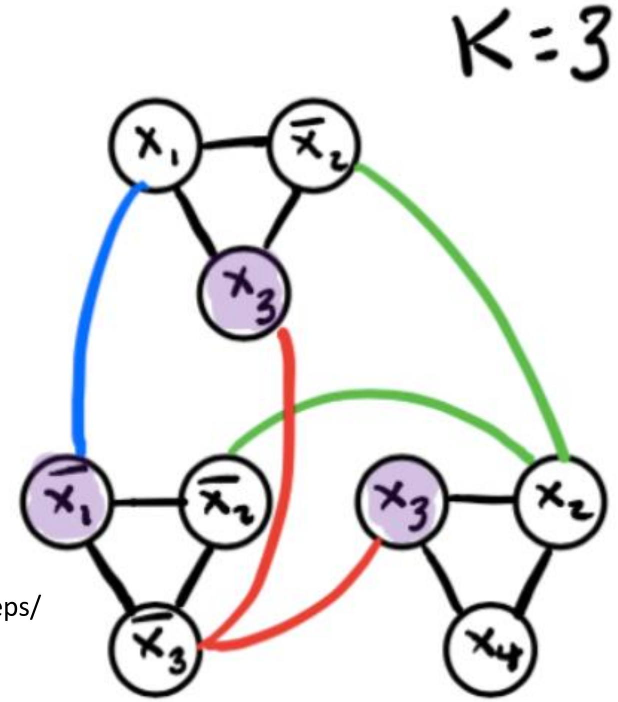


Image: https://downey.io/notes/omscs/cs6515/np-reduction-steps/

Problems with millions of variables can be solved

SAT solvers use logical algorithms + some own machine-learning techniques.

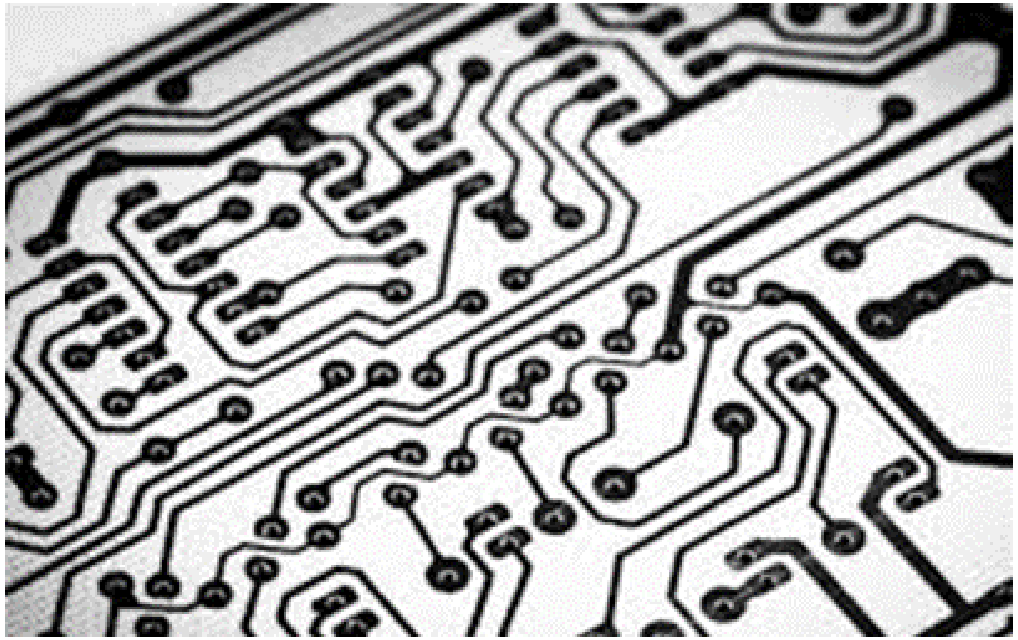
Combination with Rule-Based Knowledge Representation:

→ Answer Set Programming   DLV (UNICAL) or CLASP/Potassco (Univ. Potsdam)

More Recent Progress in Symbolic AI

Successful Applications of Constraint & SAT-Solving (examples)

Chip design and verification



Railroad network safety design

(e.g., at Siemens Austria)

More Recent Progress in Symbolic AI

Advanced Knowledge Representation and Knowledge Graphs

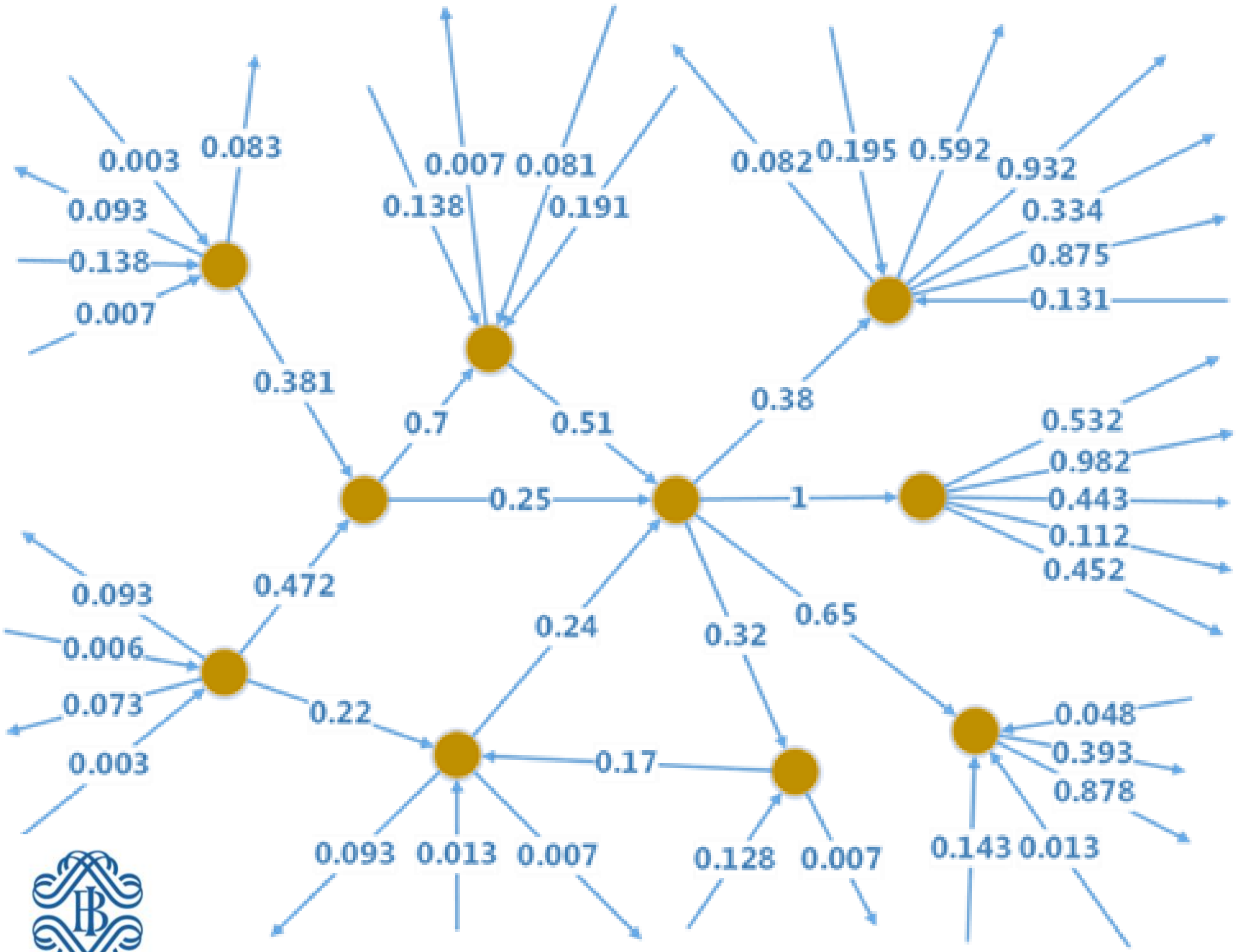
Modern KGs = Highly expressive logical rule languages
+ efficient evaluation algorithms over Big Data

Company x controls company y if x =y, or if x controls a set of companies that jointly hold over 50% of y

In Vadalog:

→ controls(x,x)

Controls(x,y) & owns(y,z,w) &
v=msum(w,⟨y⟩) & v>0.5 → controls(x,z)



BANCA D'ITALIA

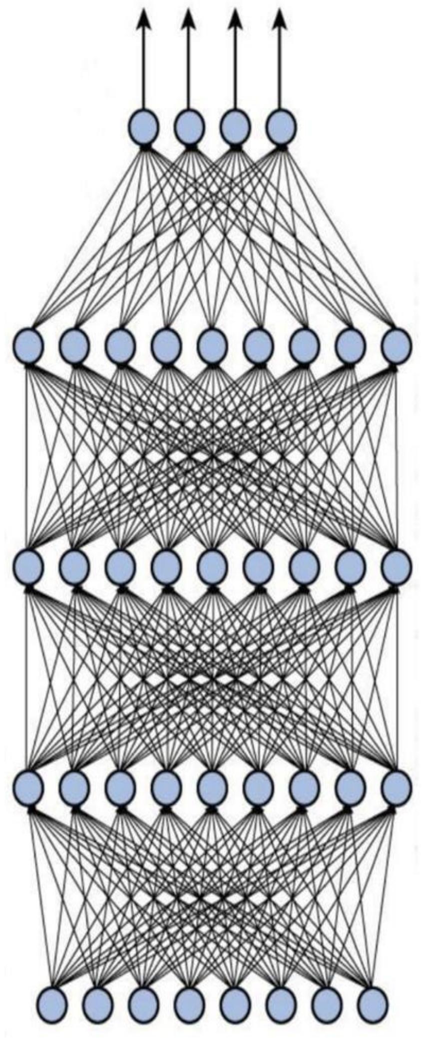
ECB Company Ownership Graph

More Recent Progress in Subsymbolic AI /ML

Deep learning: Improved multilayer perceptrons that learn from Big data.
(AlphaGo: 13 layers, GPT-3: 196 layers)

Refined Versions:

- Deep **reinforcement learning**: Environment interaction/exploration with reward maximization
- **Recurrent neural networks**: Can learn from sequences of data. Well-suited for tasks such as speech recognition, machine translation, and game playing. Enhanced by Long Short-Term Memory (LSTM).
- **Convolutional neural networks:** Can learn spatial relationships in images. Well-suited for object Recognition.



What Machine Learning Does Really Well

Pattern recognition and classification
e.g. image recognition, which also animals and babies can do

Face recognition
(positive+negative reward)

Image Source: https://loonylabs.org/2015/01/23/baby-talk/

Learning to avoid touching nettles,
explorative reinforcement learning (negative reward)

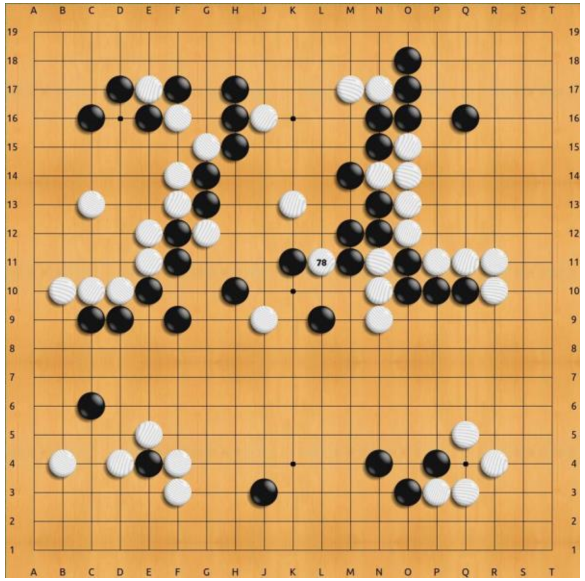
Image Source: https://ahdinnaeken.wordpress.com/2012/07/01/macaesops-fables-18-the-boy-and-the-nettles/

What Machine Learning Does Really Well

Pattern recognition and classification
e.g. image recognition, which also animals and babies can do

Avoiding obstacles. A baby explores a room and learns by itself that it cannot pass through solid matter. (*Reinforcement learning*). Similarly, robots may learn to avoid obstacles.

Game playing. Example: Go. Deep Mind/Google' s AlphaGo successor Alphazero is currently the world's top Go player. Deep learning + Tree search methods. *Classification* of board configurations and moves.



Source: https://en.wikipedia.org/wiki/File:Lee-sedol-alphago-divine-move.jpg

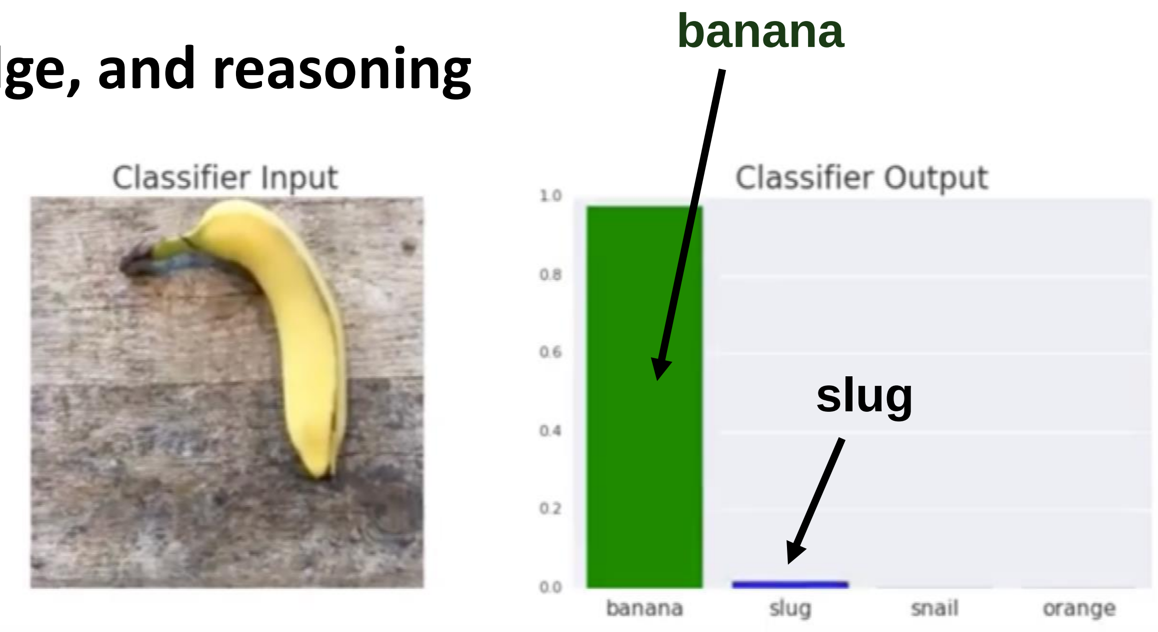
What Machine Learning Alone Cannot Do

Using transferable knowledge, and reasoning

Chinese businesswoman accused of jaywalking after AI camera spots her face on an advert
(from Telegraph, 25 Nov. 2018)

What Machine Learning Alone Cannot Do

Using transferable knowledge, and reasoning



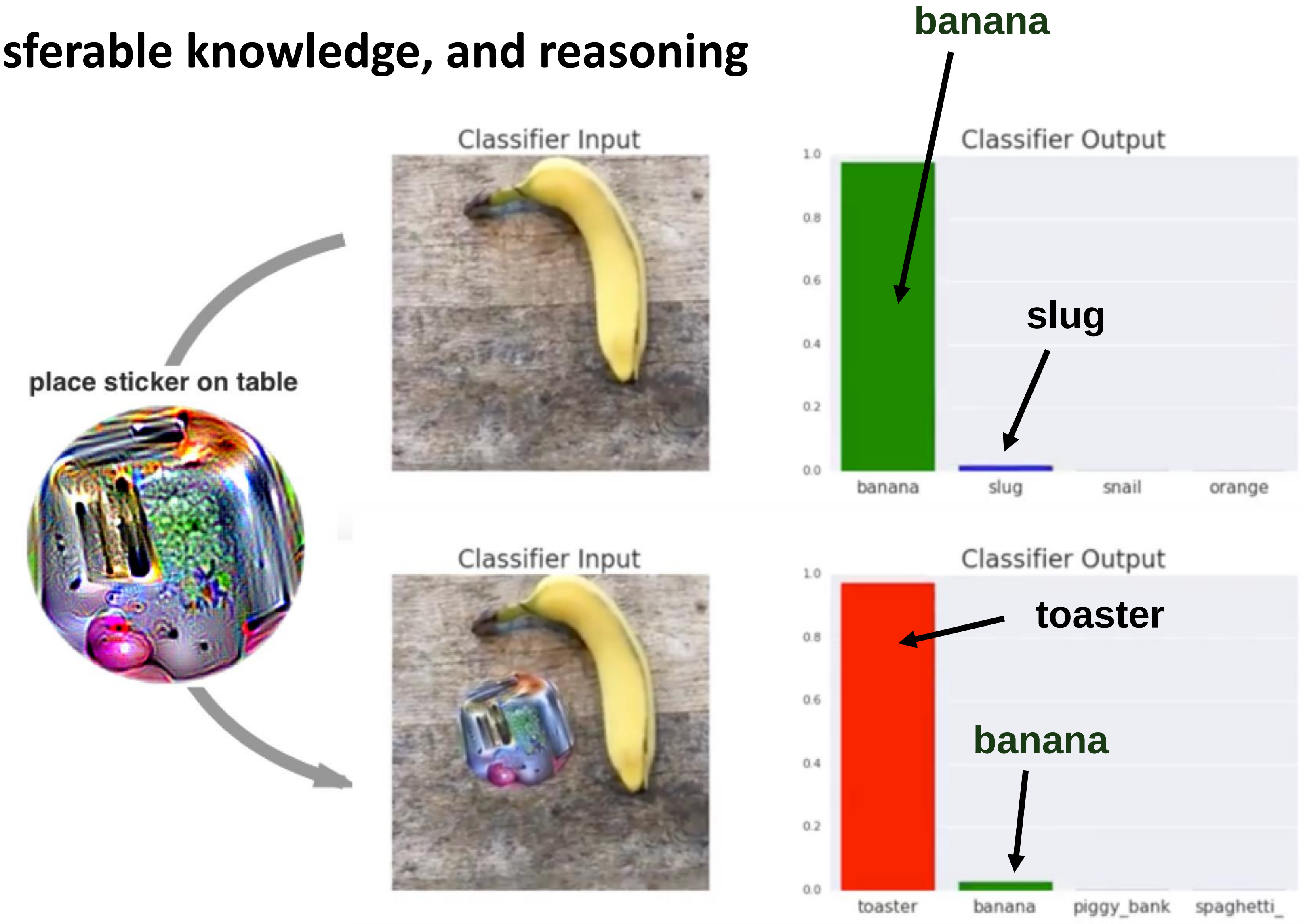
Deep Learning can be fooled due to lack of world knowledge and common sense!

https://www.youtube.com/watch?v=e9IAu4lT9w8&feature=youtu.be

Tom B. Brown, Dandelion Mané, Aurko Roy, Martín Abadi, Justin Gilmer (Google Inc.), NIPS 2017

What Machine Learning Alone Cannot Do

Using transferable knowledge, and reasoning



Tom B. Brown, Dandelion Mané, Aurko Roy, Martín Abadi, Justin Gilmer (Google Inc.), NIPS 2017

What Machine Learning Alone Cannot Do

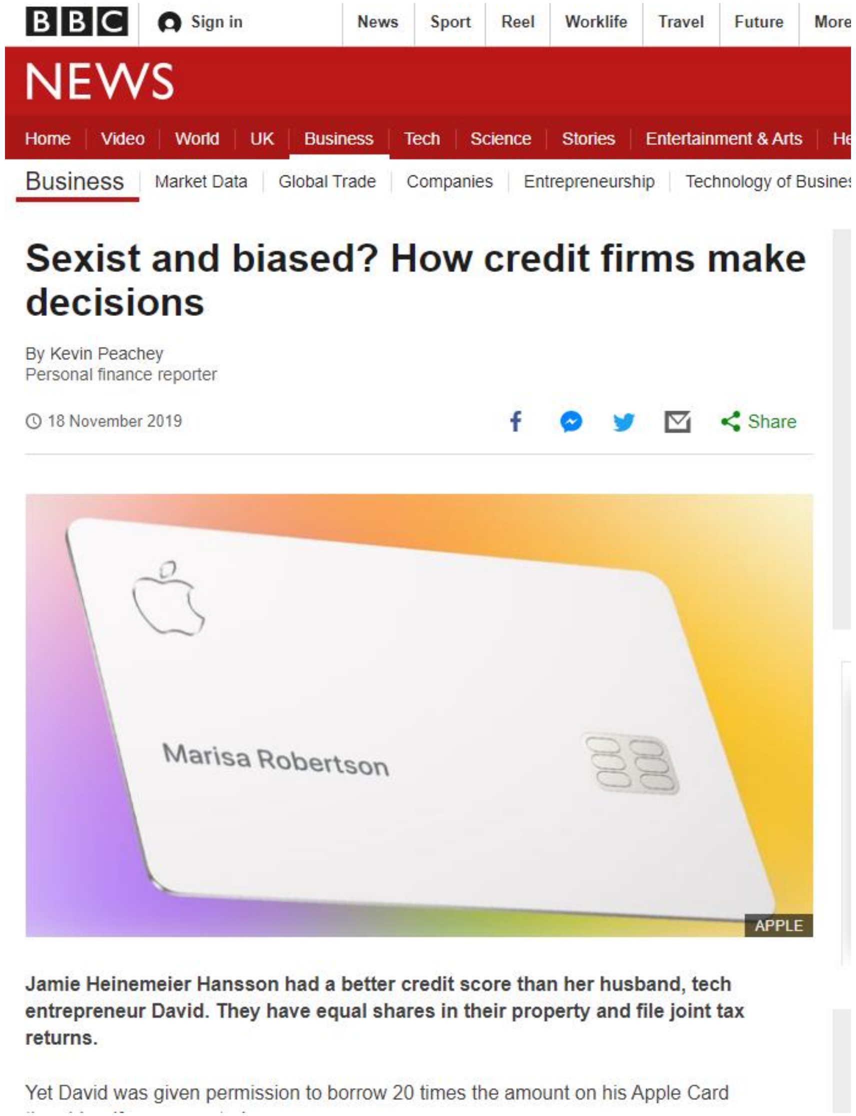
Using transferable knowledge, and reasoning

Unfair Credit Rating

Jamie Heinemeier Hansson had a better credit score than her husband, tech entrepreneur David. They have equal shares in their property and file joint tax returns. Yet David was given permission to borrow 20 times the amount on his Apple Card than his wife was granted.

Apple:*"Gender not input to algorithm"*

Rachel Thomas (Data Ethics) *: "Even if race and gender are not inputs to your algorithm, it can still be biased on these factors,"*


BBC Sign in News Sport Reel Worklife Travel Future More

NEWS

Home Video World UK Business Tech Science Stories Entertainment & Arts

Business Market Data Global Trade Companies Entrepreneurship Technology of Busine

Sexist and biased? How credit firms make decisions

By Kevin Peachey
Personal finance reporter

18 November 2019

Share

Jamie Heinemeier Hansson had a better credit score than her husband, tech entrepreneur David. They have equal shares in their property and file joint tax returns.

Yet David was given permission to borrow 20 times the amount on his Apple Card

What Machine Learning Alone Cannot Do

Using transferable knowledge, and reasoning

My Own Creditworthiness

up to £3,000

£2,500

£4,000

up to EUR 3,500

up to EUR 4,000

£500

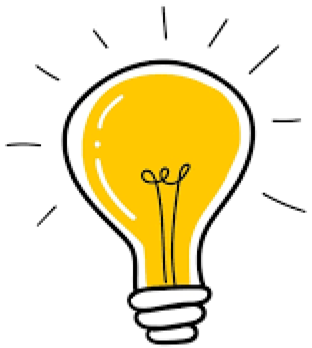
£ 3,000

£ 4,500

EUR 4,300

What Machine Learning Alone Cannot Do

Using transferable knowledge, and reasoning



My Own Creditworthiness

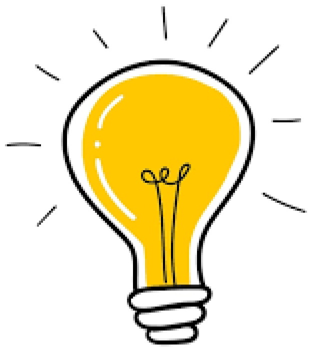
A machine-learning program has “reasonably” learned:

> ***People who live in a joint household with someone who does not pay their bills are likely to fail repaying their own debts.***

This ethically questionable rule was applied to <u>wrong data</u>.

What Machine Learning Alone Cannot Do

Using transferable knowledge, and reasoning



My Own Creditworthiness

A machine-learning program has “reasonably” learned:

People who live in a joint household with someone who does not pay their bills are likely to fail repaying their own debts.

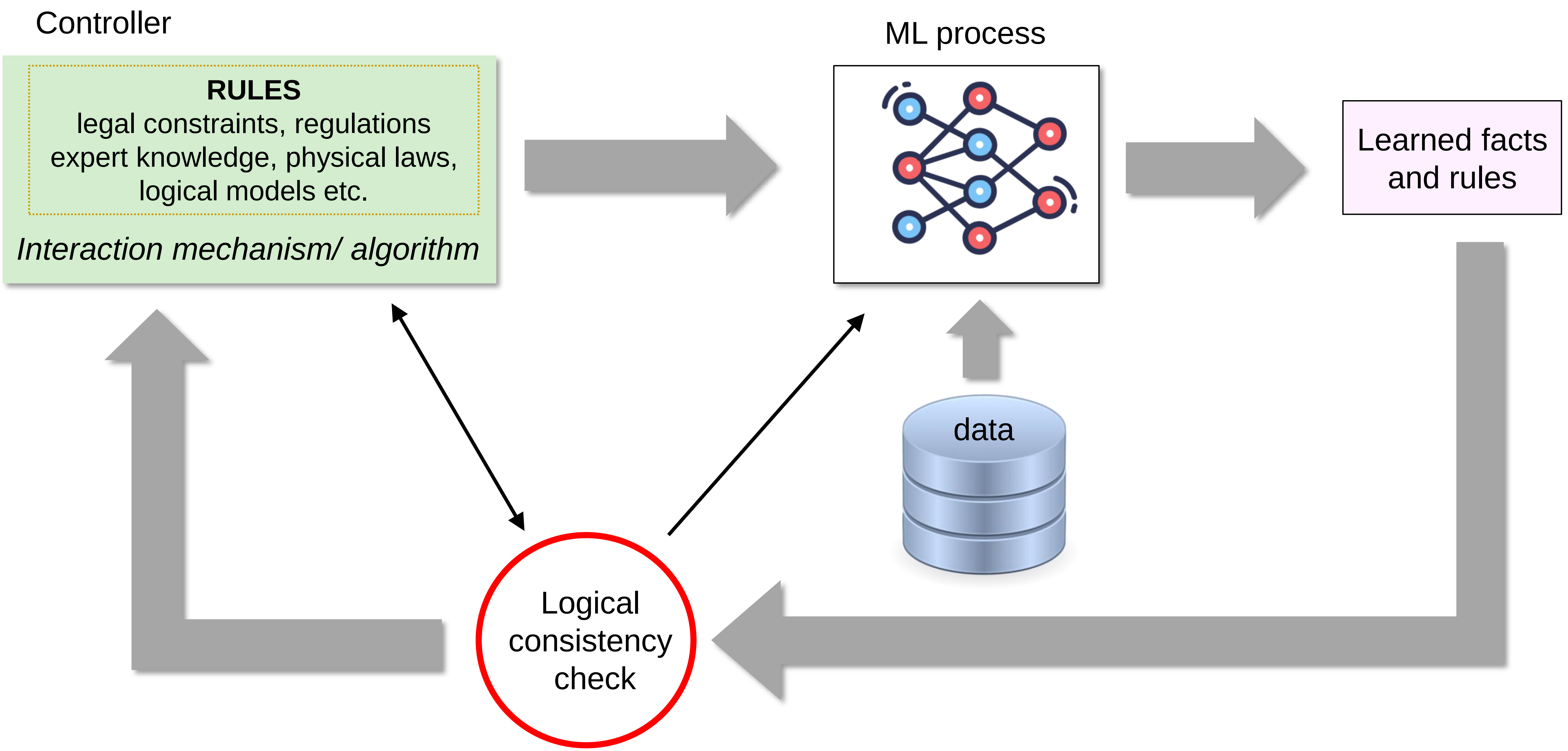
This ethically questionable rule was applied to wrong data.

A **human credit rating expert** would instead use of the rule:

If property owners move into their recently bought one-family property, then the previous occupiers have most likely moved out.

This rule can be used to update the database before applying machine learning.

Desideratum: A Logical Control Theory for ML



Desideratum: Not easy to achieve!

<u>Rules can be complicated</u> and reasoning complex

- **Classical and non-monotonic negation:**

  ¬weather(london,Date,rain) & sunday(Date) → park-concert(hydepark,Date)

  person(X) & not guilty(X) → presumed_innocent(x)

- **Probabilistic facts & rule**s (possibly from ML, statistics, or data mining):

  0.8: wheather(london,22-12-2019,heavy-snow)

  0.3: wheather(london, Date,heavy-snow) & flight(F,Date,london)→delay(F,Date)

- **Disjunction:** staff(x) → consultant(x) ∨ employee(x)

- **Existential Rule:** machine(M,t640,Room) →∃P Plug(P,Room) & HighCurrent(P)

- **Recursion, arithmetic, aggregate functions, etc:**

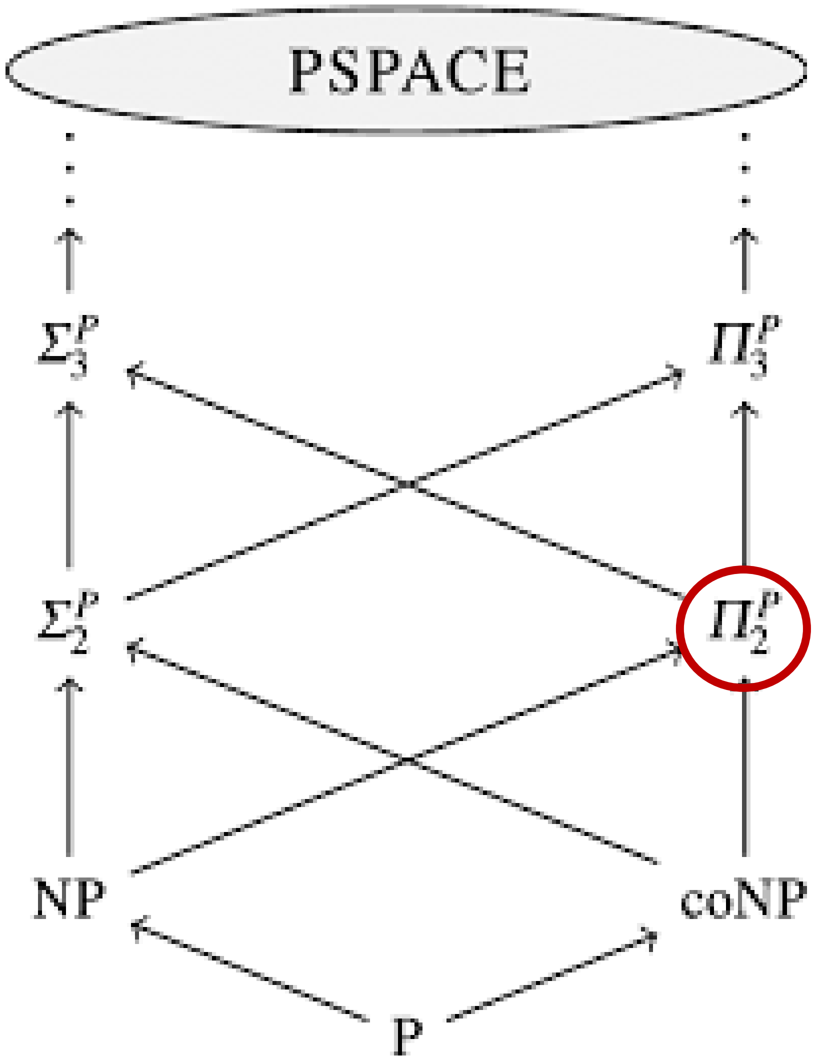
Company x controls company y if x =y, or if x controls a set of companies that jointly hold over 50% of y

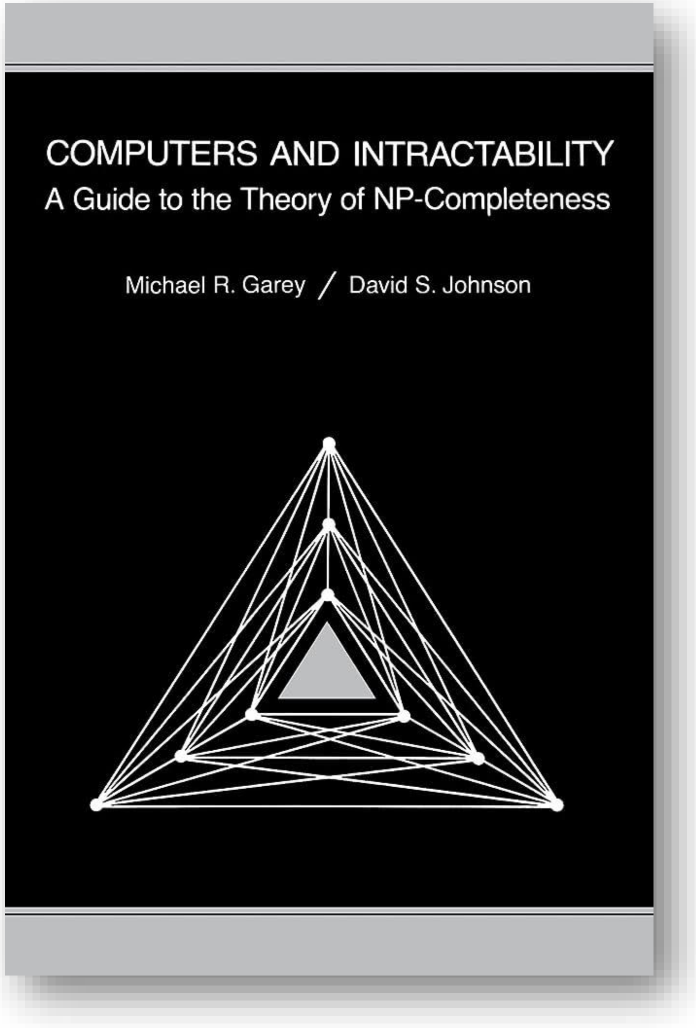
→ controls(x,x)

controls(x,y) & own(y,z,w) & v=msum(w,⟨y⟩) & v>0.5 → controls(x,z)

Desideratum: Not easy to achieve!

Rules can be complicated and <u>reasoning complex</u>





Large Language Models (LLMs)
An AI Revolution

LLMs: Based on 3 main innovations

1

Distributional "Word Embedding" semantics of Natural Language

Previous approaches to semantics (selection):

- ***Model-theoretic logical semantics*** -
  Wittgenstein, North , Russell, Whitehead, Tarski, (Chomsky), Montague,…
- ***Cognitive & ontological Semantics***
  Cognitive Metaphors: Lakoff; Frame semantics: Charles Fillmore, …
- **Statistical semantics**, "a word is characterized by the company it keeps" J.R. Firth (1957).
  Zellig Harris; ***Latent semantic analysis (LSA):*** Scott Deerwester, Susan Dumais,
  George Furnas, Richard Harshman, Thomas Landauer, and many others

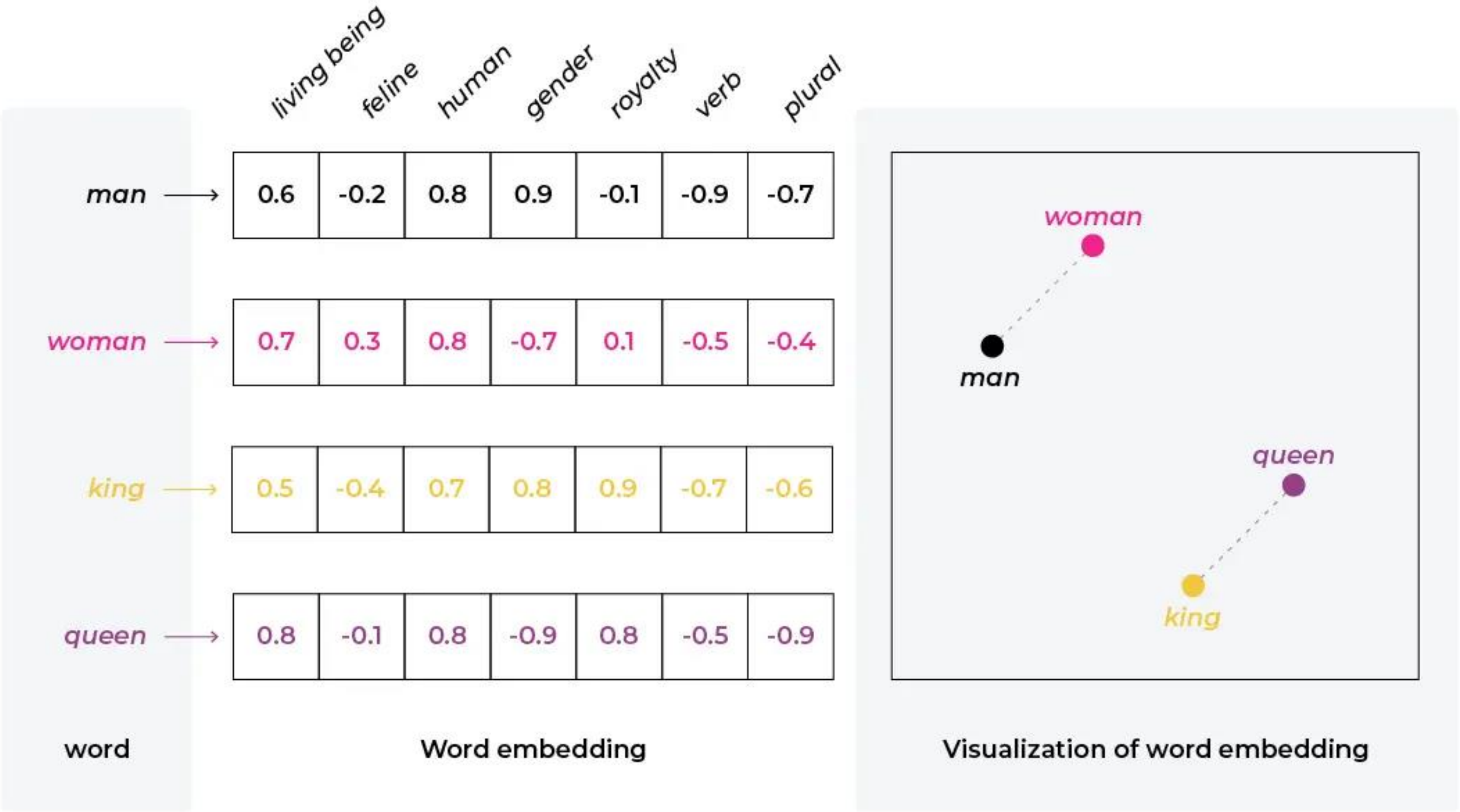
---

Modern Word Embeddings: Google Deep Mind, Bengio, et al.
word2vec, gloVe; BERT ELMo, GPT3,…

LLMs: Based on 3 main innovations

1

Distributional “Word Embedding” semantics of Natural Language



Picture source: https://arize.com/blog-course/embeddings-meaning-examples-and-how-to-compute/

LLMs: Based on 3 main innovations

2

Transformer architecture

training
compression

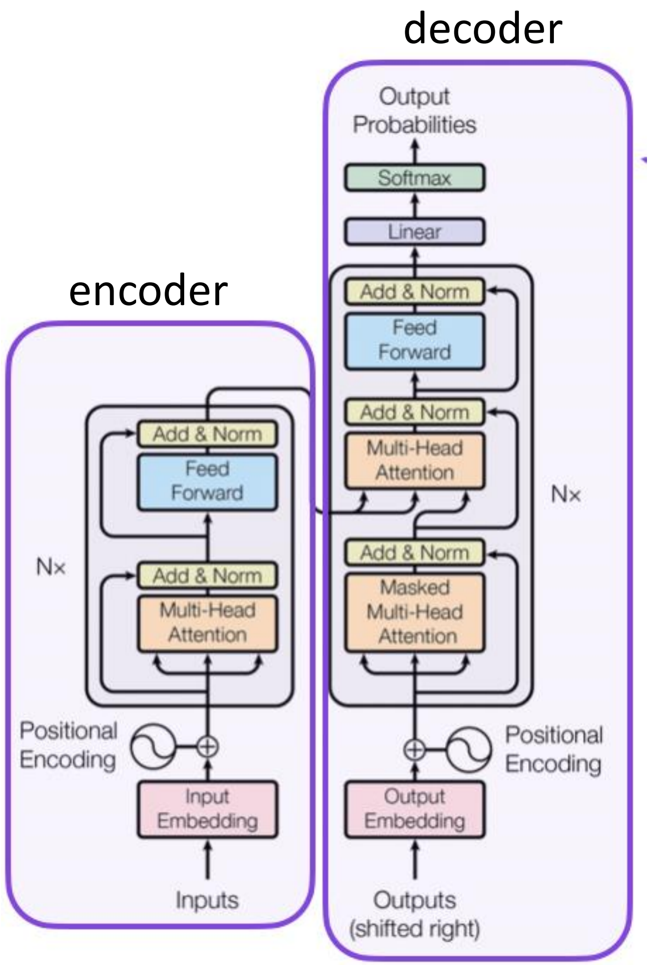


Figure 1: The Transformer - model architecture.

LLMs: Based on 3 main innovations

3

Attention
Disambiguation of word meanings based on context.

- *The worker **refuses** to work without proper pay.*
- *The streets were lined with piles of **refuses**.*

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

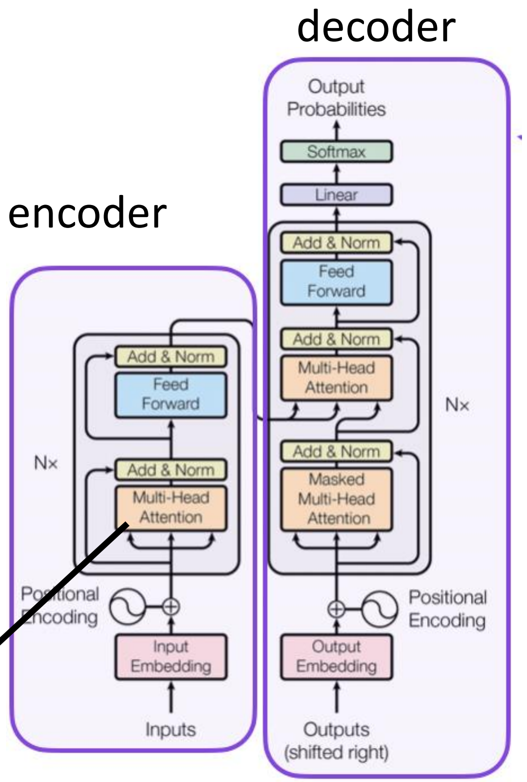


Figure 1: The Transformer - model architecture.

Attention Layer

The most powerful LLMs

GPT 3.5 (OpenAI)

GPT 4 (OpenAI)

LLAMA-2

online access to Web

BING-Chat (OpenAI+Microsoft)

BARD (Google)

Pippo Rossi è un alunno che disturba in continuazione in classe. Anche i suoi colleghi di banco non lo supportano più. E' già stato rimproverato tante volte senza risultato. L'insegnante, la dottoressa Rosa Invernizzi, vuole scrivere una lettera al padre di Pippo, dicendogli con delicatezza ma anche con determinazione che Pippo deve cambiare il suo atteggiamento, altrimenti rischia l'espulsione. Proponi una lettera di 10 righe circa.

Gentile Sig. Rossi,

Mi rivolgo a Lei con rispetto e preoccupazione riguardo al comportamento di suo figlio, Pippo, nelle ore di lezione. Nonostante i numerosi tentativi di correggere il suo atteggiamento in classe, la situazione non ha mostrato miglioramenti.

Anche i suoi compagni di banco hanno manifestato crescente disagio a causa delle continue interruzioni. È essenziale che Pippo comprenda la gravità della situazione e la necessità di un immediato cambiamento.

Le chiedo, per il bene di suo figlio e della classe, di intervenire affinché Pippo modifichi il suo comportamento. Se le azioni distruttive continuassero, saremmo, con rammarico, costretti a considerare l'espulsione come ultima soluzione.

Distinti saluti, Dott.ssa Rosa Invernizzi

Nell'Università della Calabria si apre un nuovo corso di laurea "Medicina e Chirurgia TD" (TD = Tecnologie Digitali). Oggi si festeggia l'apertura con il ministro presente. Fai un poema umoristico di 8 righe per l'occasione. Fai delle belle rime!

All'Unical un corso si fa notare,
Medicina digitale, pronta a iniziare!
Con mouse e bisturi, in perfetta armonia,
Bit e byte si fondono con la chirurgia.

Nel digitale il cuore fa un "post",
Il ministro brinda e pronuncia un toast.
In sala operatoria o su Zoom,
La medicina in Calabria ha trovato il suo boom!

Major Problem with LLMs: Truthfulness

- LLMs hallucinate by inventing facts or presenting them inappropriately.

- LLMs are biased and may propagate unfair statements.

Reasons why LLMs may hallucinate or otherwise make mistakes

- Information Compression, loss & mistaken "reconstruction"
- Adoption of wrong Web content
- Missing domains or topics in training data/ training corpora
- Wrong ordering of results
- .....

Information Compression, loss & mistaken "reconstruction"

List 5 publications by Georg Gottlob on query rewriting and data exchange

Gottlob, G., & Schwentick, T. (2004). Rewriting ontological queries into small nonrecursive datalog programs. KR 2004.

Barceló, P., Gottlob, G., & Pichler, R. (2014). Logical foundations of relational data exchange. ACM SIGLOG News, 1(1), 2-13.

Dalvi, N., Gottlob, G., & Kikot, S. (2012). Query rewriting for Horn-SHIQ plus rules. In Proceedings of the Twenty-Sixth AAAI Conference on AI

Gottlob, G., Orsi, G., & Pieris, A. (2014). Query rewriting and optimization for ontological databases. ACM Trans. on Database Systems 25.

Calì, A., Gottlob, G., & Kifer, M. (2008). Taming the infinite chase: Query answering under expressive relational constraints. In KR 2008

Correct Citations:

Logical foundations of relational data exchange, Pablo Barceló ACM SIGMOD Record, 2009

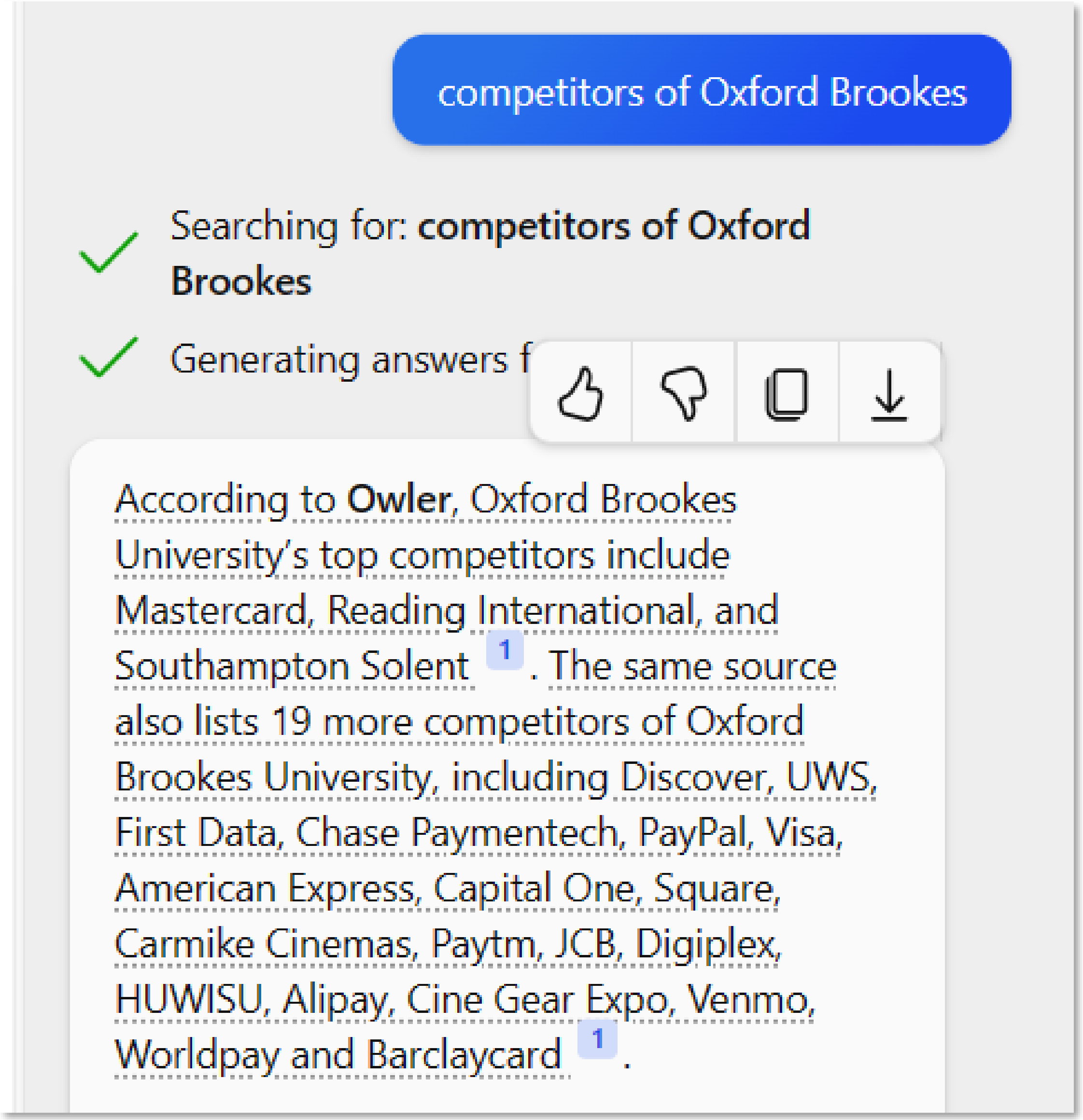
Query rewriting for Horn-SHIQ plus rules, T Eiter, M Ortiz, M Simkus, TK Tran & G. Xiao- Proc. AAAI 2012

40% wrong

How embarrassing!

Adoption of wrong Web content

- Previous example: GPT is not connected to the Web
- It can get worse: Wrong Web-data might be adopted



Bing-Chat

Missing domains or topics in training data/training corpora

Imaginary scenario: LLM-translator trained on real estate & household concepts.
Has access to basic dictionary.

Does it also work in other contexts?

Training example (generated statistically from training corpus):

chest → cassone

cassa

cassetta

Training example:

flat → casa

appartamento

studio

Missing domains or topics in training data/training corpora

Training example:

house → palazzo

casa

appartamento

etc…

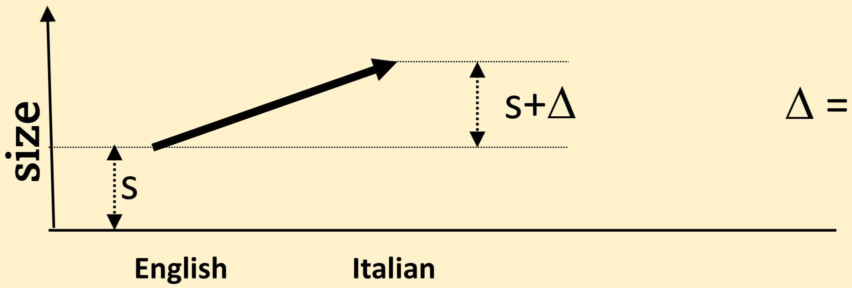
Apply system to translate concepts in real-estate domain

palace → reggia

palazzo

casa

The system has learnt the following “translation vector” in the dimension “size”:



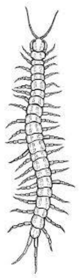
Δ = *one order of magnitude*

Missing domains or topics in training data/training corpora

Let us now use this trained system to translate English→Italian in other domains.

Let's see the outcome…

Domain of Entomology



centopede

millepiedi

centopiedi

diecipiedi

Academic Domain

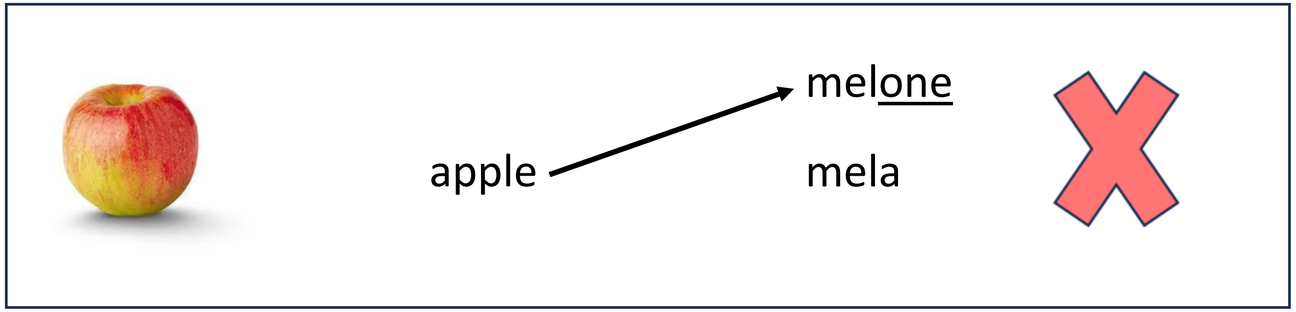
Doctor

Master

Dottore

Missing domains or topics in training data/training corpora

Domain "food"



Wrong ordering of results

List joint competitors of Birkbeck and Oxford Brookes regarding reputation

Regarding reputation, some universities in the UK that could be considered joint competitors of Birkbeck College and Oxford Brookes University are:

1. University of Warwick 101-125
2. University of Bristol 81-90
3. University of Sheffield 126-150
4. University of Leeds 101-125
5. Univ. of Nottingham 151-175

MASSIVE GAP!

Birkbeck College 351–400

Oxford Brookes 601-800

CHAT2DATA
BY UNLIMIDATA.AI

***Chat2Data* Project**

Objective: Get high-quality data from LLMs by using LLMs

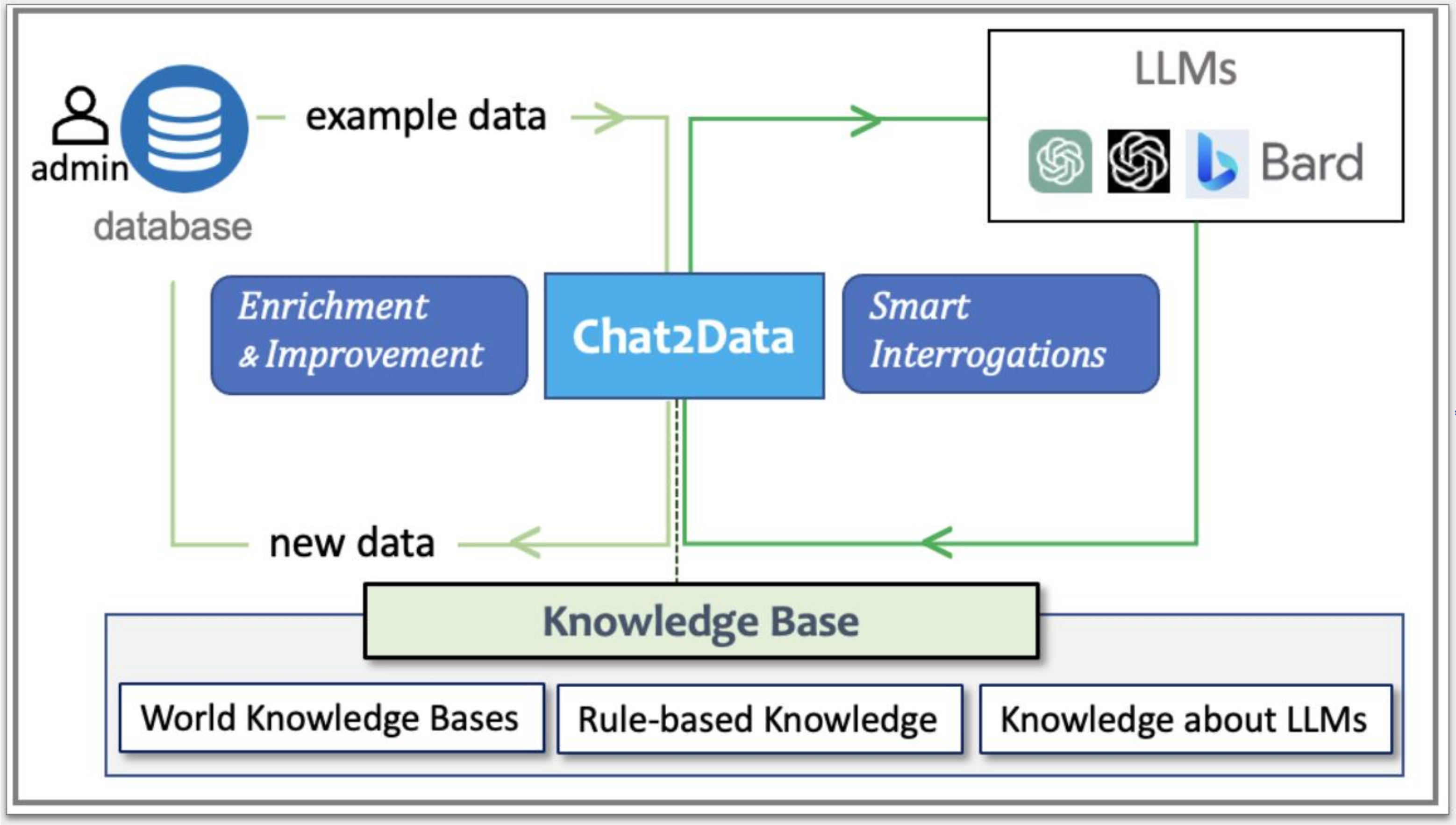
Context: Fully automated Database verification, correction, and enrichment

Benchmark: Owler crowd-sourced DB about 16M companies, products, competitors, etc. The database has a verified part of ~150k tuples.

Main ideas underlying the Chat2Data system:

- access structural info (relations, attributes) from DB
- take sample records from DB
- obtain full understanding of semantics of DB via prompting
- verify/correct/update existing DB records by ***smart interrogation***
- enrich DB using ***smart interrogation***

Chat2Data Project



Internal Document Repository

- generation of new datasets
- enrichment of existing databases
- verification of data records in a DB
- update of data & null values

Input Data (R)	
A	B
Doctolib	Jameda
Doctolib	Arzt-direkt
Doctolib	doctap
FoodCheri	Nestor
FoodCheri	famileat
FoodCheri	dinicatering
Oxford Brookes Univ.	Coventry University
Oxford Brookes Univ.	University of Oxford
Tracktor.fr	MachineryZone
Tracktor.fr	europe-agriculture.com
Zenjob	jobmensa.de
Zenjob	totaljobs
Nightjet	Ryanair
Blender	*null*
Zwilling	*null*
Kamps Bakery	*null*
LC Waikiki	*null*
…	…

Chat2Data

Feature I:

Database Verfication:

Find incorrect tuples in the input database.

For example,

Doctap operates in the UK only, but Doctolib operates in Italy, Germany, and France.

University of Oxford and Oxford Brookes University have very different rankings.

Input Data (R)	
A	B
Doctolib	Jameda
Doctolib	Arzt-direkt
~~Doctolib~~	~~doctap~~
FoodCheri	Nestor
FoodCheri	famileat
~~FoodCheri~~	~~dinicatering~~
Oxford Brookes Univ.	Coventry University
~~Oxford Brookes Univ.~~	~~University of Oxford~~
Tracktor.fr	MachineryZone
~~Tracktor.fr~~	~~europe-agriculture.com~~
Zenjob	jobmensa.de
~~Zenjob~~	~~totaljobs~~
Nightjet	Ryanair
Blender	*null*
Zwilling	*null*
Kamps Bakery	*null*
LC Waikiki	*null*
…	…

Input Data (R)	
A	B
Doctolib	Jameda
Doctolib	Arzt-direkt
FoodCheri	Nestor
FoodCheri	famileat
Oxford Brookes Univ.	Coventry University
Tracktor.fr	MachineryZone
Zenjob	jobmensa.de
Nightjet	Ryanair
Blender	*null*
Zwilling	*null*
Kamps Bakery	*null*
LC Waikiki	*null*
...	...

Chat2Data

Feature I:
Database Enrichment:

1. Find more data records

Example:
New tuples are highlighted in green grids.

Input Data (R)	
A	B
Doctolib	Jameda
Doctolib	Arzt-direkt
Doctolib	Mondocteur
Doctolib	Arzttermine.de
FoodCheri	Nestor
FoodCheri	Seazon
FoodCheri	La Belle Assiette
Oxford Brookes Univ.	Coventry University
Oxford Brookes Univ.	Univ. of Surrey
Oxford Brookes Univ.	Univ. of Kent
Tracktor.fr	MachineryZone
Tracktor.fr	Loxam
Tracktor.fr	Kiloutou
Zenjob	jobmensa.de
Zenjob	totaljobs
Zenjob	JobToday

Input Data (R)	
A	B
...	...
Blender	*null*
Zwilling	*null*
Kamps Bakery	*null*
LC Waikiki	*null*
...	...

Chat2Data

Feature II:

Database Enrichment:

2. Find missing values in the input database

Input Data (R)	
A	B
...	...
Blender	**Bforartists**
Blender	**Fusion 360**
Zwilling	**Calphalon**
Zwilling	**Cuisinart**
Kamps Bakery	**Müller**
Kamps Bakery	**Mälzer**
LC Waikiki	**Bershka**
…	…

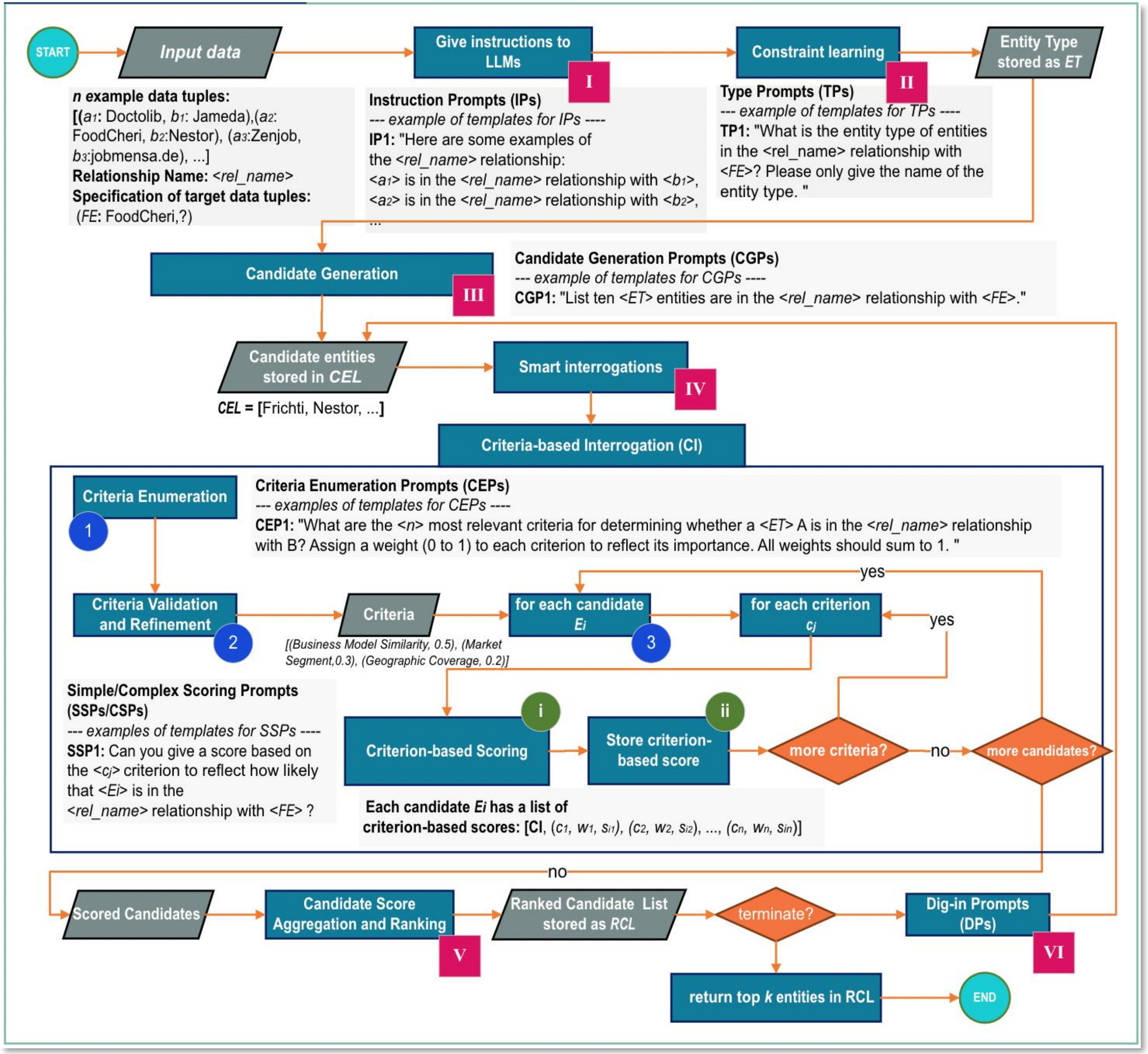
Input Data (R)		GPT-4		Bing Chat		Mini-Chat2Data
A	B	new B (zero-shot)	new B (few-shot)	new B (zero-shot)	new B (few-shot)	new B
Doctolib	Jameda	Zocdoc, Practo, KRY / LIVI, RDV Médicaux, Credihealth	Zocdoc, Healthgrades, Practo, Bookimed, Doctor On Demand	DocPlanner, Doctena, Jameda, Keldoc, Qare	Solutionreach,WebPT, QGenda, Axxess Home Health, Mend	DocPlanner, Mondocteur, Arzttermine.de, Doctoralia, Doxter
FoodCheri	Nestor	Frichti, Deliveroo, Uber Eats, Just Eat, Glovo	Frichti, Deliveroo, Uber Eats, Just Eat, Glovo	Frichti, PopChef, Foodette, Deliveroo, City Pantry	Frichti, Foodette, PopChef	Frichti, PopChef, Seazon, La Belle Assiette, Foodette
Oxford Brookes Univ.	Coventry University	Univ. of Oxford, Univ. of Reading, Univ. of Bath, Univ. of Southampton, UWE Bristol	Univ. of Oxford, Univ. of Reading, Univ. of Hertfordshire, Univ. of Northampton, Univ of Southampton	Mastercard, Reading International, Southampton Solent	University of Oxford, University of Cambridge, University of Warwick	Univ. of Surrey, Bournemouth Univ., Univ. of Reading, Univ. of Kent, Univ of Southampton
Tracktor.fr	MachineryZone	La Poste, DHL, UPS, FedEx, 17track	Eu-Construction-Equip Agriaffaires, RentalYard, Mascus, IronPlanet	Villas et Maisons de France	Villas et Maisons de France	Loxam, Kiloutou Mateco, Ramirent, Riwal
Zenjob	jobmensa.de	Instawork, Coople, Wonolo, Rota, Gig	StudentJob, JobTeaser, Fiverr, LinkedIn, Indeed	clickworker, Streetspotr,Syft, WorkGenius, inploi	clickworker, Streetspotr,Syft, WorkGenius, inploi	Coople, Jobandtalent,Syft, JobToday, L1nda
Nightjet	Ryanair	DB, Thello, Trenitalia, RZD, SNCF	Eurostar, FlixBus, DB, EasyJet, TGV Lyria	Oebb, DB, SBB, Nsinternational,Thalys	Oebb,DB, SBB, trainline.eu,Thalys	Euronight, SNCF, SBB, Thello, DB
Blender	*null*	Maya, 3ds Max, Cinema 4D, ZBrush, Houdini	Maya, 3ds Max, Cinema 4D, Houdini, ZBrush	Canva, Glorify, SketchUp, Cinema 4D, Modo	Maya, Cinema 4D, 3ds Max, Wings 3D Revit	Bforartists, Fusion 360, Godot Engine, FreeCAD, SketchUp
Zwilling	*null*	Wüsthof, Victorinox, Global, Shun, MAC	Wüsthof, Victorinox, Henckels International, Global, Shun	Wüsthof, Victorinox, Henckels	zwillingonline.com, Henkels, Wüsthof, wer-zu-wem.de, Kai	Henkels, Wüsthof, WMF, Calphalon, Cuisinart
Kamps Bakery	*null*	BackWerk, Le Crobag, Ditsch, Starbucks, Dunkin' Donuts	Bäcker Wiedemann, Hofpfisterei, BackWerk, Dunkin' Donuts, Starbucks	Wback, von Allworden, Backer Gortz, Kamps GmbH, Kamps	BackWerk, Ditsch, Kamps, Kamps Backstube, Kamps GmbH	Müller, Mälzer, Steinecke, Göing, Le Crobag
LC Waikiki	*null*	Zara, H&M, Mango, Uniqlo, Primark	Zara, H&M, Mango, Uniqlo, Primark	Defacto, Koton, H&M, Morhipo, Glami.com.tr	Defacto, Trendyol, Koton, Boyner, Morhipo	Koton, H&M, DeFacto, C&A , Bershka

Why (advanced) prompting is not enough?

Blue: new direct competitors of company A
Black: existing direct competitors of company A
Orange: indirect competitors of company A
Red: incorrect competitors of company A

Chat2Data

Example Workflow




Grazie per l’attenzione…
…per la splendida accoglienza in Calabria!