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...


**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 **Symbolic AI**

Constraint Processing & SAT solving (NP-hard)

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)
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:

\[ \rightarrow \text{controls}(x, x) \]
\[ \text{Controls}(x, y) \& \text{owns}(y, z, w) \& v = \text{msum}(w, \langle y \rangle) \& v > 0.5 \rightarrow \text{controls}(x, z) \]
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).
Pattern recognition and classification
e.g. image recognition, which also animals and babies can do

Face recognition
(positive + negative reward)

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

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

What Machine Learning Alone Cannot Do

Using transferable knowledge, and reasoning

https://www.youtube.com/watch?v=e9lAu4lT9w8&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
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,”
What Machine Learning Alone Cannot Do

Using transferable knowledge, and reasoning

My Own Creditworthiness
Using transferable knowledge, and reasoning

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.
What Machine Learning Alone Cannot Do

Using transferable knowledge, and reasoning

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

Controller

RULES
legal constraints, regulations
expert knowledge, physical laws,
logical models etc.

Interaction mechanism/ algorithm

ML process

Learned facts and rules

Logical consistency check

data
Desideratum: Not easy to achieve!

Rules can be complicated and reasoning complex

- Classical and non-monotonic negation:
  \[
  \neg \text{weather(london,\textit{Date},\textit{rain})} \land \text{\textit{sunday(Date)}} \rightarrow \text{park-concert(hydepark,\textit{Date})} \\
  \text{person(X)} \land \neg \text{\textit{guilty(X)}} \rightarrow \text{presumed_innocent(x)}
  \]

- Probabilistic facts & rules (possibly from ML, statistics, or data mining):
  \[
  0.8: \text{weather(london,22-12-2019,\textit{heavy-snow})} \\
  0.3: \text{weather(london,\textit{Date},\textit{heavy-snow})} \land \text{flight(F,\textit{Date},london)} \rightarrow \text{delay(F,\textit{Date})}
  \]

- Disjunction: 
  \[
  \text{staff(x)} \rightarrow \text{consultant(x) } \lor \text{employee(x)}
  \]

- Existential Rule:
  \[
  \text{machine(M,t640,Room)} \rightarrow \exists P \text{ Plug(P,Room)} \land \text{HighCurrent(P)}
  \]

- Recursion, arithmetic, aggregate functions, etc:
  \[
  \text{Company x controls company y if x =y, or if} \\
  \text{x controls a set of companies that jointly hold over 50% of y}
  \]
  \[
  \text{controls(x,y)} \land \text{own(y,z,w)} \land v=\text{msum(w,\langle y\rangle)} \land v>0.5 \rightarrow \text{controls(x,z)}
  \]
Desideratum: Not easy to achieve!
Rules can be complicated and reasoning complex
Large Language Models (LLMs)
An AI Revolution
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

  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

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**

<table>
<thead>
<tr>
<th>Name</th>
<th>Institute</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ashish Vaswani*</td>
<td>Google Brain</td>
<td><a href="mailto:avaswani@google.com">avaswani@google.com</a></td>
</tr>
<tr>
<td>Noam Shazeer*</td>
<td>Google Brain</td>
<td><a href="mailto:noam@google.com">noam@google.com</a></td>
</tr>
<tr>
<td>Niki Parmar*</td>
<td>Google Research</td>
<td><a href="mailto:nkip@google.com">nkip@google.com</a></td>
</tr>
<tr>
<td>Jakob Uszkoreit*</td>
<td>Google Research</td>
<td><a href="mailto:usz@google.com">usz@google.com</a></td>
</tr>
<tr>
<td>Llion Jones*</td>
<td>Google Research</td>
<td><a href="mailto:llion@google.com">llion@google.com</a></td>
</tr>
<tr>
<td>Aidan N. Gomez†</td>
<td>University of Toronto</td>
<td><a href="mailto:aidan@cs.toronto.edu">aidan@cs.toronto.edu</a></td>
</tr>
<tr>
<td>Lukasz Kaiser*</td>
<td>Google Brain</td>
<td><a href="mailto:lukaszkaiser@google.com">lukaszkaiser@google.com</a></td>
</tr>
<tr>
<td>Illia Polosukhin†</td>
<td></td>
<td><a href="mailto:illia.polosukhin@gmail.com">illia.polosukhin@gmail.com</a></td>
</tr>
</tbody>
</table>
The most powerful LLMs

- GPT 3.5 (OpenAI)
- GPT 4 (OpenAI)
- LLAMA-2
- BING-Chat (OpenAI+Microsoft)
- BARD (Google)

online access to Web
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!
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
- .....

**Major Problem with LLMs: Truthfulness**
List 5 publications by Georg Gottlob on query rewriting and data exchange


40% wrong
How embarrassing!

Correct Citations:

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

Adoption of wrong Web content

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

<table>
<thead>
<tr>
<th>chest</th>
<th>cassone</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cassa</td>
</tr>
<tr>
<td></td>
<td>cassetta</td>
</tr>
</tbody>
</table>

Training example:

<table>
<thead>
<tr>
<th>flat</th>
<th>casa</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>appartamento</td>
</tr>
<tr>
<td></td>
<td>studio</td>
</tr>
</tbody>
</table>
Missing domains or topics in training data/training corpora

Training example:

- House: English = ‘house’; Italian = ‘casa’
- Palace: English = ‘palace’; Italian = ‘reggia’; ‘palazzo’
- Apartment: English = ‘apartment’; Italian = ‘appartamento’

Apply system to translate concepts in real-estate domain

- The system has learnt the following “translation vector” in the dimension “size”:
  \[ s \approx \Delta \]  
  \( \Delta = \text{one order of magnitude} \)
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
Domain “food”

apple  melone  mela

Missing domains or topics in training data/training corpora
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

Birkbeck College 351–400
Oxford Brookes 601-800

MASSIVE GAP!
**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

- generation of new datasets
- enrichment of existing databases
- verification of data records in a DB
- update of data & null values
### Feature I: Database Verification:

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.
### Feature I: Database Enrichment:

1. Find more data records

**Example:**
New tuples are highlighted in green grids.
**Feature II:**

**Database Enrichment:**

2. Find missing values in the input database
<table>
<thead>
<tr>
<th>Input Data (R)</th>
<th>GPT-4</th>
<th>Bing Chat</th>
<th>Mini-Chat2Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>А</td>
<td>B</td>
<td>А</td>
<td>B</td>
</tr>
<tr>
<td>Doctolib</td>
<td>Jameda</td>
<td>Zocdoc, Practo, KRY / LIVI, RDV Médicaux, Credihealth</td>
<td>DocPlanner, Doctena, Jameda, Keldoc, Qare</td>
</tr>
<tr>
<td>FoodCheri</td>
<td>Nestor</td>
<td>Frichti, Deliveroo, Uber Eats, Just Eat, Glovo</td>
<td>Frichti, PopChef, Foodette, Deliveroo, City Pantry</td>
</tr>
<tr>
<td>Oxford</td>
<td>Coventry University</td>
<td>Univ. of Oxford, Univ. of Reading, Univ. of Bath, Univ. of Southampton, UWE Bristol</td>
<td>Mastercard, Reading International, Southampton Solent</td>
</tr>
<tr>
<td>Tracktor.fr</td>
<td>MachineryZone</td>
<td>La Poste, DHL, UPS, FedEx, 17track</td>
<td>Villas et Maisons de France</td>
</tr>
<tr>
<td>Zenjob</td>
<td>jobmensa.de</td>
<td>Instawork, Coople, Wonolo, Rota, Gig</td>
<td>StudentJob, JobTeaser, Fiverr, LinkedIn, Indeed</td>
</tr>
<tr>
<td>Nightjet</td>
<td>Ryanair</td>
<td>DB, Thello, Trentitalia, RZD, SNCF</td>
<td>Oebb, DB, SBB, Nsinternational,Thalys</td>
</tr>
<tr>
<td>Blender</td>
<td>null</td>
<td>Maya, 3ds Max, Cinema 4D, ZBrush, Houdini</td>
<td>Canva, Glorify, SketchUp, Cinema 4D, Modo</td>
</tr>
<tr>
<td>Kamps Bakery</td>
<td>null</td>
<td>BackWerk, Le Crobag, Ditsch, Starbucks, Dunkin' Donuts</td>
<td>BACK, von Allworden, Backer Gortz, Kamps GmbH, Kamps GmbH</td>
</tr>
</tbody>
</table>

**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!