Dangers of AI...

We discussed the first five of these:

- bias
- deepfakes
- easy foolability
- lack of explainability
- unchecked power
- automation
- surveillance, loss of privacy
- ...

A*G*I - take a deep breath!!!

Aaaaaaaaaaaaaaaaaagh.

AGI: Artificial *General* Intelligence, ie. a human-like intelligent system.

AGI has been the 'next gen AI', since the birth of AI (in 1950) :) :(

So, where is it?

PS: a lot of what follows are my opinion...

Too much HYPE, too early on!

A series of early 'wins' made researchers get carried away. Here is a nice set of writeups on this.

Tests for AI

Turing Test: https://en.wikipedia.org/wiki/Turing_test

Loebner Prize: https://en.wikipedia.org/wiki/Loebner_Prize

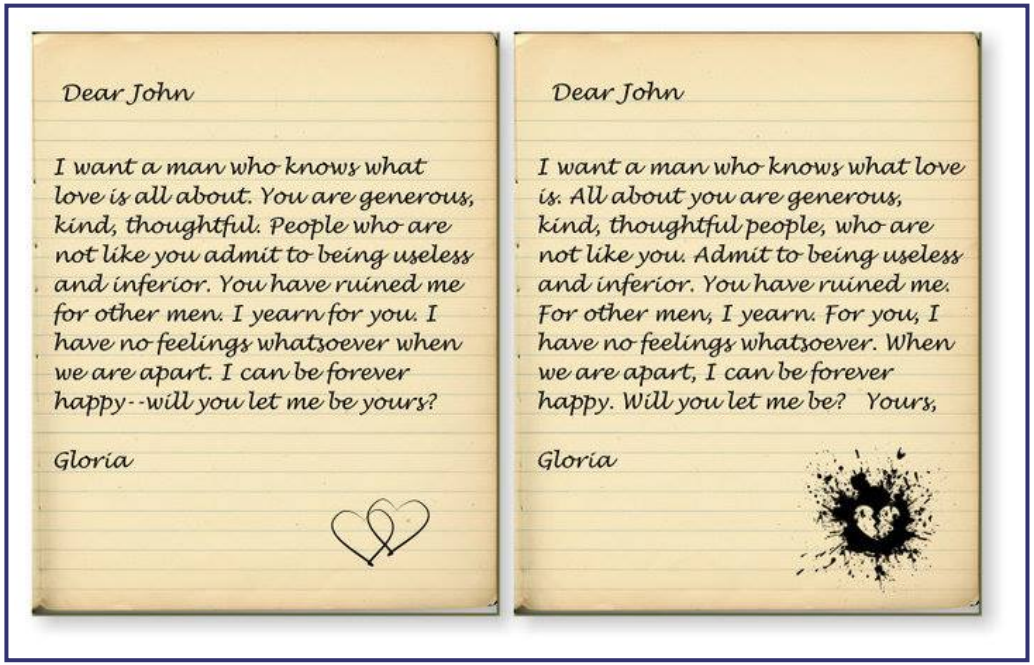
https://developer.amazon.com/alexaprize

Aside: Searle's Chinese Room Experiment: https://en.wikipedia.org/wiki/Chinese_room

Most of these tests/prizes are irrelevant when it comes to developing AGI!

Language - is NOT about words!

No amount of NLP [alone] is going to fix this.



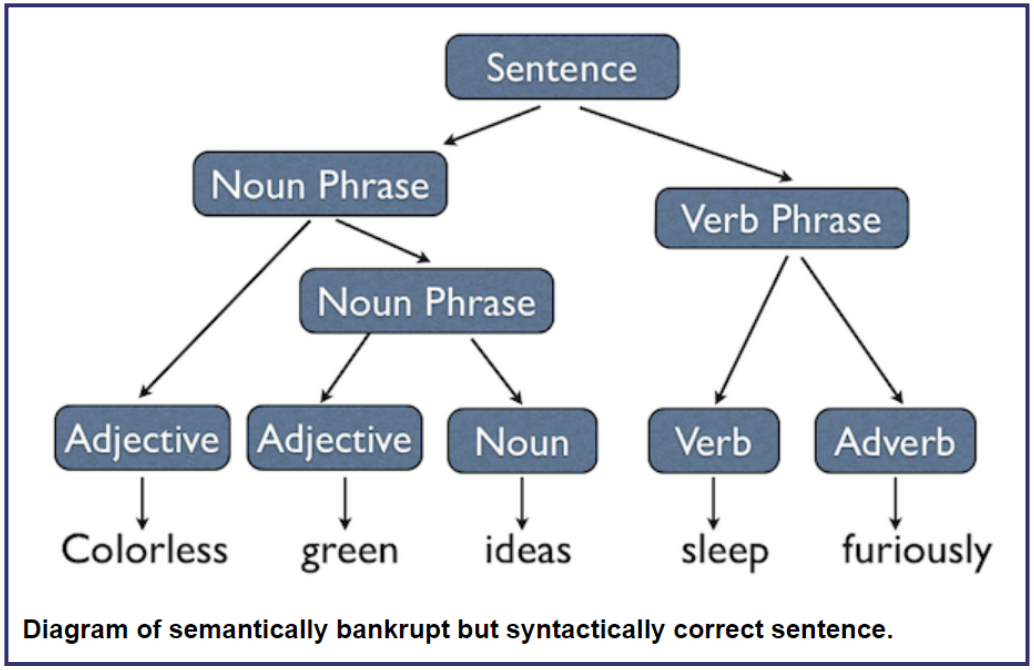
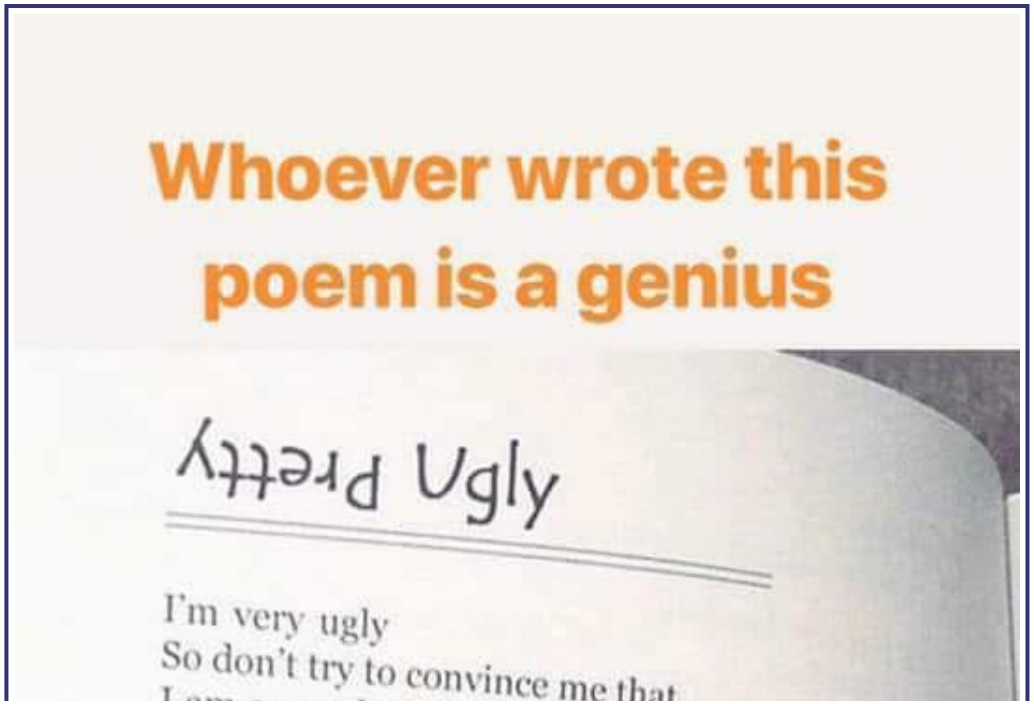
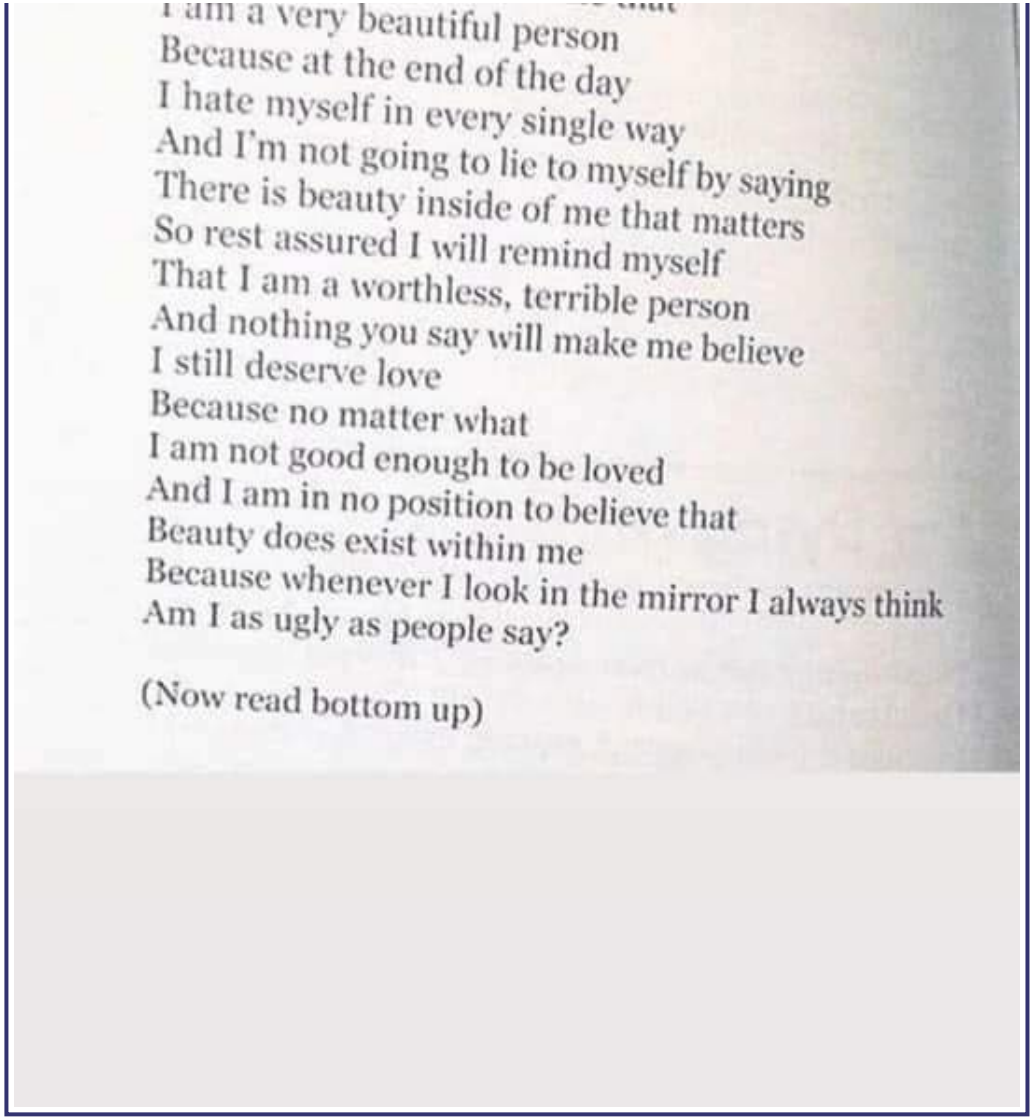


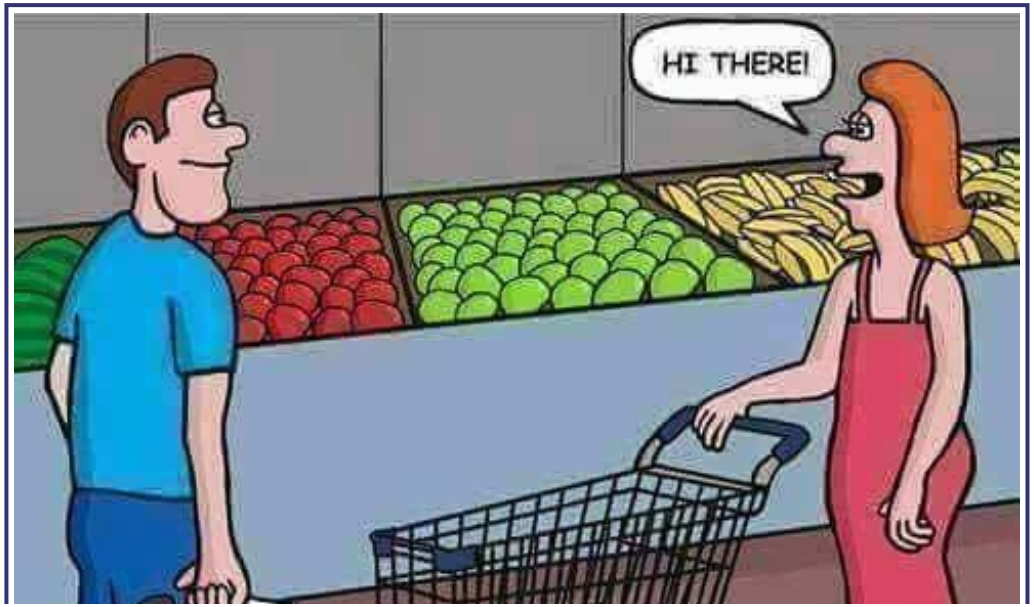
Diagram of semantically bankrupt but syntactically correct sentence.




I am a very beautiful person
Because at the end of the day
I hate myself in every single way
And I'm not going to lie to myself by saying
There is beauty inside of me that matters
So rest assured I will remind myself
That I am a worthless, terrible person
And nothing you say will make me believe
I still deserve love
Because no matter what
I am not good enough to be loved
And I am in no position to believe that
Beauty does exist within me
Because whenever I look in the mirror I always think
Am I as ugly as people say?

(Now read bottom up)


I WAS WONDERING WHY THE
FRISBEE WAS GETTING BIGGER
... THEN IT HIT ME
imgflip.com


HI THERE!


The Grog
www.GrogComics.com
DO YOU KNOW ME?
I THINK YOU'RE THE FATHER OF ONE OF MY KIDS!
ARE YOU THE STRIPPER FROM THE BACHELOR PARTY THAT I MADE LOVE TO ON THE POOL TABLE WITH ALL MY BUDDIES WATCHING WHILE YOUR PARTNER WHIPPED MY BUTT WITH WET CELERY?
NO, I'M YOUR SON'S TEACHER.

Images - are NOT about pixels

So, what is it, then?

The world has structure, phenomena, behaviors, cause-and-effect... WE LEARN THESE, THROUGH CONTINUING EXPERIENCE!!

We use a mix of induction (generalize -> likely conclusion), deduction (reason -> guaranteed conclusion), abduction (hypothesize -> best-guess conclusion) in our daily lives...

ML is not 'learning'!

Backprop is not the only game in town!

People, Part 1

Misled/clueless/hype-y:

- Stephen Hawking
- Bill Gates
- Elon Musk
- Ray Kurzweil
- ...

People, Part 2

Researchers and groups, exploring various aspects of AGI:

- Paul Rosembloom (SOAR, Sigma)
- Arthur Toga
- Daniel Dennett
- Gary Marcus
- Alison Gopnik
- Demis Hassabis
- Judea Pearl
- Leslie Valiant
- Joscha Bach
- Yann LeCun
- NELL
- BabyX
- Allen Institute
- Josh Tenenbaum, also https://www.youtube.com/watch?v=7ROelYvo8f0&t=5085s
- Jeff Hawkins
- Kenneth Stanley (Neuroevolution)

- Karl Friston
- Giulio Tononi
- OpenAI
- DeepMind
- OpenCog
- ...

The above, spans the gamut - studying the brain, building brain-like hardware, brain-like software, musing philosophically, studying infants, pondering cause-and-effect, quantifying consciousness, proposing mind architectures, etc, etc. NO ONE HAS THE FULL PICTURE (a blueprint for creating AGIs) YET!

In addition, Marvin Minsky's 'Society of Mind' is an intriguing collection of ideas related to intelligence.

My take on all this

- brain structure MATTERS!
- embodiment MATTERS! Embodied cognition is how nature does it
- situatedness MATTERS! The brain is in a body that lives in the world (in an environment). VR provides a controllable environment for safe, modifiable, repeatable explorations...
- memory STRUCTURE matters - how things are coded, retrieved...
- continuous, 24x7x365 experiencing - record EVERYTHING
- WE develop hunches/heuristics - from experience!
- we can learn a LOT (on building AGI) - get clues from babies, children, animals!
- an inner 'copy' of the world, manipulatable
- a richly connected 'graph'
- emotions
- attention
- consciousness, thoughts, feelings, self, free will...
- can result in authentic creativity, incl art etc
- develop a comprehensive cognitive architecture
- testable items (on an AGI architecture) - conscience/morality/empathy, disorders, diseases, language, learning modes...

- SDCs, robots... ALWAYS 2nd class - NO INTRINSIC MOTIVATION
- summary: experience (v) -> experiencer ('self', ie. 'I') -> experience (n), and '4E' (embodiment, embeddedness, extendedness, enactivism)

My proposal - an agent that:

- is embodied (with senses) [and is virtual ("for now") - NOT ideal at all!]
- has brain structure and functions modeled after ours
- can interact, explore, learn and grow CONTINUOUSLY, incrementally
- starts out with hardwired basic behaviors
- has an attention mechanism
- is capable of emotions in addition to thoughts

My two word definition of intelligence [across all species, plus artificial]: