AI cannot be recognised as an inventor, US rules

The US Patent Office rules only humans can legally be inventors

An artificial intelligence system has been refused the right to two patents in the US, after a ruling only "natural persons" could be inventors.
Value Misalignment

• Cognitive systems may not understand the real intention of the training data
  • Lack of common sense knowledge
  • Limited data that doesn’t represent reality
  • Values/beliefs are hard to define
Examples of value misalignment

- An Eurisko game-playing agent that got more points by falsely inserting its name as the creator of high-value items.

- A Lego staking system that flips the block instead of lifting, since lifting encouragement is implemented by rewarding the z-coordinate of the bottom face of the block.

- A sorting program that always outputs an empty list, since it is considered a sorted list by the evaluation metric.

- A game-playing agent that kills itself at the end of level 1 to avoid losing in level 2.

- A robot hand that pretends to grasp an object by moving between the camera and the object.

- A game-playing agent that pauses the game indefinitely to avoid losing.

For more, see https://t.co/mAGUf3quFQ
Transparency

The General Data Protection Regulation (GDPR)

- Limits to decision-making based solely on automated processing and profiling (Art.22)
- Right to be provided with meaningful information about the logic involved in the decision (Art.13 (2) f. and 15 (1) h)

Paul Nemitz, Principal Advisor, European Commission Talk at IBM Research, Yorktown Heights, May, 4, 2018
Explanation understanding depends on the listener

End Users
- Who: Physicians, judges, loan officers, teacher evaluators
- Why: trust/confidence, insights(?)

Regulatory Bodies
- Who: EU (GDPR), NYC Council, US Gov’t, etc
- Why: ensure fairness for constituents

AI System builders, stakeholders
- Who: data scientists, developers, prod mgrs
- Why: ensure/improve performance

Affected Users
- Who: Patients, accused, loan applicants, teachers
- Why: understanding of factors

Must match the *complexity capability* of the consumer
Must match the *domain knowledge* of the consumer
Cognitive systems reviews

• What is the intended use of the results?

• What algorithms or techniques does the service implement?

• Which datasets was the service trained/tested on?

• Describe the testing methodology and results.

• How was the model trained, and were any steps taken to protect the privacy or confidentiality of the training data?

• Are you aware of possible examples of bias, ethical issues, or safety risks as a result of using the service?

• Does the service implement and perform any fairness checks detection and bias mitigation?

• What is the expected performance on data with different distributions?

• Was the service checked for robustness against adversarial attacks?

• When was the service last updated?

• Recommended uses. Not-recommended uses.
Trusting data sources and cognitive system producers

Data management  how and for what purpose is my data used?

Tool transparency  How to access the qualities of the provided tools and models?
Trusting policy makers

Personal data protection Is my data safe?

Privacy Do I have to give up privacy for better services?

Accountability Who is liable if things go wrong?

Impact on Jobs Can we retrain / relocate workers displaced by cognitive systems?

AI weaponization How will cognitive systems be used in conflicts between companies or between nations?