System, Human, and Cybernetic Aspects of AI: Methodological Thoughts

Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso
500 W. University, El Paso, Texas 79968, USA, vladik@utep.edu

1 What Is Not AI and What Is AI

A routine computational problem is when:

- we have a well-defined problem, and
- to solve this problem, we apply a well-justified algorithm to objective (measurement-based) data.

In this form, solving this problem is routine, no intelligence is needed.

In many real-life situations, however:

- the problem is not well-defined (e.g., how to formalize "comfortable" ?),
- some data comes not from measurements but from expert estimates, and
- to process this data, we use expert-proposed heuristic (= not justified) methods.

In such non-routine cases, we need to use intelligence.

When computers are used to solve such intelligence-related problems, this becomes AI.

Comment. Of course, this is a traditional definition of AI, before the current trend to identify AI with machine learning.

2 From This Viewpoint, AI Is All About Human-Related Aspects

In line with the above desccription, to solve a problem that requires intelligence, we need:

- to formulate a human-motivated problem in precise terms,
- to be able to describe (usually imprecise) expert knowledge and opinions in precise terms, and

- to extract (and explain) useful heuristics from how we solve problems – this is where, e.g., neural networks (NN) came from.

All this is about human-related aspects, and our Systems, and Cybernetics Society (SMCS) is the only IEEE society that has human in its title.

Of course, other societies are also active and successful in AI: IEEE Computational Intelligence Society (CIS) is clearly the leader in this.

Our potential strength is in the borderline between computations and humans. Once a precise formulation is attained, heuristic extracted – then other societies often have a technical edge. But there is always a possibility to use human connection to improve heuristics.

For example, NNs came from simulation of human brain. However, an effective idea of deep NN also came from the fact that in a human brain, we have many neural layers.

This is our potential focus – coming up with new human-motivated methods of:

- formalizing human problems and expert knowledge,
- solving problems, and
- describing solutions in human-understandable form (this is part of what is usually meant by eXplainable AI – XAI).

Once a method is formalized, it becomes mathematics, computer science – in which other societies are probably more skilled. But we can be an intermediate link between humans and AI systems.

For this to be successful, we need a constant collaboration with other societies. This way, we treat all aspects of a problem as a system – and not just get bogged down in one specific aspect.

3 What About S and C of SMC?

Since we mentioned systems, what about S and C of SMC? So, far, this was all about humans (M of SMC).

Systems means that we consider whole objects, not just individual aspects or sub-objects. This means that we view everything from the viewpoint the system as a whole – whose improvement is usually our main objective. E.g., we want to improve traffic throughput – and not just improve a single road segment. This can clearly help with human-related aspects of AI.

Cybernetics was originally defined as finding commonalities between humans (and living creatures) and machines:

- this can help better understand humans by using our experience with machines – and thus, better extract useful ideas, and
- this can help better analyze engineered systems – by using analogy with well-analyzed living systems.

This can also help with human-related aspects of AI.

4 What Are Aspects of AI in Which Our Help Is Most Needed

Modern deep-learning-based AI systems are spectacular, but they are not perfect. One of the main problems is that AI systems often produce wrong results: they identify cats as dogs, LLMs hallucinate, etc.

Moreover, AI systems' estimates of their own accuracy and reliability are way off. When a cat is recognized as a dog, the system claims 99% confidence. What we need is:

- to better gauge the system's accuracy,
- to better detect when AI is inaccurate, and
- to improve the accuracy of AI systems.

Also, AI system need much more examples to train than humans use: how can we improve AI in this regard? Again, analysis of how we humans learn may help.

5 How Can We Better Gauge AI Systems Accuracy and Reliability: Uncertainty in AI

A naive optimistic viewpoint is that AI systems can, in principle, be trained to predict accuracy of AI systems. This was tried, but the results are not good. This fact is easy to understand:

- good AI results come from training on millions of data, but
- accuracy comes not from individual data but from groups; and
- from millions of data point we can only extract thousands of groups – not enough for good estimation.

We need better methods for gauging AI's accuracy. Hopefully, analysis of how we humans make decisions can help.

6 How to Better Detect AI's Mistakes: Need for XAI

Human decision-makers are also not perfect. However, when a doctor makes a recommendation, we can ask for an explanation. If the explanation is not convincing, our confidence in this recommendation is lower. To apply the same idea to AI-based decision making, we need eXplainable AI (XAI).

European Union now *requires* XAI, but Shap and other "XAI" methods are not exactly explanations that we want. We need XAI that will be explainable

not only according to a formal definition, but also explainable from the commonsense viewpoint. For this, we need to analyze what we humans mean by understanding.

7 How to Make AI Systems More Accurate and More Reliable: Incorporating Expert Knowledge

In the previous sections, we talked about better gauging of the system's accuracy and reliability. Gauging is good, but ideally, we should make AI more reliable.

AI mistakes are often the ones that humans would never do. So, to avoid them, we need to incorporate human knowledge into AI systems.

In the past, this was the main AI direction: logic programming, fuzzy, etc. We humans use this knowledge when neurons in our brains' work. We need to teach AI to do it.