We are missing something, collectively

"On the highway towards Human-Level AI, Large Language Model is an off-ramp."
Why is it fundamentally flawed?

Terence Le Huu Phuong
February 4, 2023

Introduction

This small text is a spontaneous (but hopefully constructive) discussion of Yann Le Cun's vision about the way to Human-Level AI, or Autonomous Machine Intelligence. I have quickly written these notes right after reading the paper "A Path Towards Autonomous Machine Intelligence, Version 0.9.2".

My thesis can be summarized as follows:

- Language is inherent to Learning
- Learning is a collective behavior
- Human-Level AIs will be collective or not be

At a time when everyone claims that ChatGPT is a breathtaking step to AGI, Yann Le Cun encourages (urges?) to caution and humbleness. And I think he cannot be more right.

Those notes are spontaneous thoughts, written with the feeling that we are missing something when dealing with learning. Most of the points need deeper development, but the hope is that they would be a good start and food for thoughts on what we are trying to achieve and why. They would fit in the "Discussions, limitations" section of the paper.

In the paper, Yann Le Cun develops his vision on the path toward Human-Level AI. He starts by explaining how human beings learn and how we should devise machines in a similar way, before going into the (more) technical implementation details. What ignited my curiosity is twofold: Learning is defined as an individual concept, and the "negligence" for the Language.

I was first piqued in the introduction by a seemingly benign sentence that invites "impatient readers [...] to jump directly" to the sections of interest, and regret this introduction section was not more filled out.

In my opinion, any paper or text that has the ambition of developing a path, as technical as it may be, toward Human-Level AI, must take a significant amount of time defining what a "Human-Level", "Learning", or "Intelligence" mean**.** This is not secondary. This is the Archimedean point on which everything will be built upon. It is worthless trying to develop a theory in a broken framework, because it would not be able to overtake the limitation of the framework in which it

lives in, by construction**. I therefore recommend reading the introduction of the paper with great attention,** as well as the related work and discussion sections.

I submit that the very fundamental philosophical idea on which the paper is based on is flawed by nature because of a (collective) misconception of *Learning* and more generally human interactions. This flaw comes from our individualistic and anthropocentric model of the humankind. (Please read “flawed” as “incomplete”, “imperfect”, not “wrong”).
We are focusing on the individual agent, instead of embracing the collective intelligence, “the big picture”.

To be very clear, my goal is not to discuss the paper itself, but to rather challenge our tech community. I am afraid that we might be confining ourselves in a tunnel that it may be extremely difficult to get out from.

The following text thus focuses on the Introduction of the paper. The rest of the paper (what most readers will be interested in) is not discussed.

1) What is the paper about?

For those who didn’t read the paper (yet), here is a short summary.

The human beings and animals learning scheme is based on observation and interaction that constitutes “Common Sense”, which is defined as a collection of world models that allows them to learn, predict, and plan. This idea of "world model" comes from psychology, the science and study of individual mind and behavior (the fact that “psychology” is quoted in the paper is not insignificant).

Those World Models are devised by an enormous amount of observations, and **a paradoxically small amount of interactions** with the real world. Representations of the world are built layer by layer of abstraction to form a hierarchical model.

On the contrary, Machine Learning systems need to be trained by an enormous amount of data and trials and are still very far from human skills.
For instance, A human can learn how to drive in a couple of dozens of hours when a self-driving system would require thousands of trials of reinforcement learning or a large labels dataset just to learn that driving too fast in a turn can result in a bad outcome.

As a result, the main challenge for AI is to devise paradigms and architectures that allow machines to learn Word Models in an unsupervised fashion.

The human brain contains one World Model (instead of one model per task) that is used to achieve multiple and different tasks. A “Human-Level” AI wannabe must follow this pattern.

Yann Le Cun then derives 3 main challenges for AI:

- Allowing machines to learn world representations by observation (not by interaction)
- Building this paradigm and architecture to fit gradient-based learning
- Allowing the machine to learn in a hierarchical manner, as humans do

In this context, Large Language Model are incomplete by nature because most human knowledge is experienced and not represented as text. An LLM can only approach a superficial knowledge of reality because it cannot experience and interact with the world.

2) Why I believe this conception might be fundamentally flawed

Based on the reading of the paper, one may understand why Yann Le Cun wrote: "On the highway towards Human-Level AI, Large Language Model is an off-ramp."

Humans start by gathering basic knowledge of how the world works, getting an understanding, and feeling of the physics model, light, objects etc.
Language is a complex and abstract notion that is quite high / late in the hierarchical world model. Basing a Human-Level AI on Language model would be like starting to build a house from the second floor.

And this is where, to me, the first fundamental mistake is. Language defines the word "Human" in "Human-Level" AI. It is really close to the foundations of the house.

1. Language is at the heart of Learning.

The Language is at the heart of humankind's definition and identity. It shapes our representation of the world in a **collective way**.

Other animals have the ability to develop languages. Chimpanzees, birds, or insects have some form of language. But those animals' babies are not capable of spontaneously generating new words.
What distinguishes the human from other animals is the capacity to produce new words and sentences from what we hear in our entourage. This very capacity defines the humankind.

Most people would say that a language is a tool to communicate. Communication is from Latin communicare, which means to share, to impart, to transmit. The language is a collective concept. It is how we teach and learn, how we transmit knowledge from one generation to another, how we share our individual perception of the world with our peers. And because most of us agree on what we perceive, we engrave those shared representations in new words and expressions. This specific representation of the world constitutes a culture (which etymologically means to grow, to cultivate land), a shared vision of the world.

A language is thus more than just a communication tool. A language crystallizes a certain representation of the world and the relationship with the environment, spatially and temporally. This is an analysis tool that articulates the universe in separate units and is the building block of abstract and complex concepts.

We build world models collectively, not individually.

But then, we are able to create new abstract concepts from the language only, loosening our bonds from the perception of the world, sometimes completely detaching our thoughts from it. We observe and interact with the world, then engrave our perception in languages, before going beyond this frame of thoughts by building new concepts and words on top of it, unconstrained from our first prior perception. This is the idea of George Orwell's 1984 Newspeak: by removing words, one can confine the mind in a well-defined and narrow world model, making it impossible to step out of the frame.

Language is an extremely powerful feedback loop.

As a result, building a machine that tries to learn representations of the world individually, with observations only, is vain, and would hardly reach "Human-Level".

2. A new human baby born is not a blank paper

The reason behind the belief that the language may not be that important in the process of learning may come from a specific and arbitrary (and deeply held in the western world) vision of the humankind, and the nature in general. Thus, the nature of the discord is philosophical.

This belief is based on the idea that at birth, an agent, a human being, an animal, a living being, is a blank page, that would need to be written from scratch. A human being in this philosophical model would exist on its own. Conceptually, it would be like a human randomly spawning in the world. This is a deep philosophical belief and school of thought.

A human being would be like an empty hard drive storage that would need to be filled step by step, on its own. My personal belief is that a human individual would not survive on its own, because what he would learn from its observations and interactions with the world would be insufficient.

Animals do not just learn on an individual basis, otherwise they would be no progress nor civilization possible, let alone survival. Human beings and other animals learn collectively, over time and history. This learnt information is called evolution and is stored... in DNA.

Even more, living beings do not learn on a "breed" basis, they learn collectively as a whole. Trees communicate with each other through fungi. Trees, fungi, and insects are said to be living in symbiosis. They didn't learn to live in symbiosis during their lifetime, by observing the environment. Millions of years of evolution did it.

Likewise, human being come to life with a specific DNA background that has been carried over generations. This is why we taste and smell things differently, to be able to identify what is dangerous or safe to eat, where it is safe to stand. The polecat "learnt" to give off a specific odour that most other animals would identify as foul odour. But it was born with this ability, it didn't learn it by observation or interaction (at least not the "newest" polecats).

This "DNA" learnt "latent" memory is a different kind of learning that transcend the individuals and works by reproduction. This is maybe why we need this "incomprehensibly" small amount of interactions with the world, because we already have the knowledge, embedded in our cells.

A "Human-Level" AI would need to have the capability to leverage such a latent memory.
On a more down-to-earth aspect, human beings also learn by transmission of knowledge from parents, teachers, and more generally, the community. Learning gathers the whole body, from neurons to cells, to chemical reactions. All senses are required and used.

Feelings play a central role in the process a learning. This is why the information taught by a member of family or a very close person is more likely assimilated quicker and deeper. This is why professional speakers use methods leveraging different kind of rhythms when speaking and make extensive use of storytelling. They are trying to activate emotions and feelings,
something that is hard to catch when tweaking some weights in a neural network-based machine.

Again, the interaction with the peers plays a major role in the human learning process.

Conclusion

To conclude, our individualistic, anthropocentric, and brain-based vision of Learning is likely to be flawed and incomplete.

Trying to implement a paradigm that follows this vision, framed in a "Neumann-oriented" model of a discrete decomposition of separate modules, (only) based on gradient-based learning might be vain. I have the feeling that we are forwarding our flawed and incomplete understanding of our own process of learning into the machines we are trying to build, obsessed with instantaneous and rapid progress.

We are obviously missing something, feeding millions of data and 250 watt to a machine to have the same outcome a toddler would do with 100ml of milk is silly.

If we want to reach Human-Level AI, Language must be at the heart of the architecture. We cannot build a single machine, we must build a swarm of machines that would need to build a language to interact with each other and carry that learnt knowledge along over generations.

Human-Level AI must be collective or not be.