Next Steps in AI

December 18, 2024

Pierre Haren, Olav Laudy, Eric Jensen

Summary

A lot of AI experts are commenting on the current state of AI, the path to AGI and the future of neuro-symbolic architectures. Leveraging our 40+ years of symbolic AI research and development, including our work at Causality Link, we suggest here a few of the necessary improvements of current AI technology to get closer to this elusive AGI goal.

The current chatter about the seeming plateauing of LLMs with additional computing power is promoting a growing agreement on neuro-symbolic architectures, or on the need to combine a System1 rapid "intuitive" pattern matching reasoning with a System2 "rational" explainable reasoning, aligning AI with human decision making as described by Daniel Kahneman in his book "Thinking, Fast and Slow".

We suggest here four necessary improvements to the current LLM paradigm: full explainability, the ability to forget selectively, continuous learning, and the ability to go back in time. We believe these foundational capabilities are required to build impressive System1 and System2 interactions, including the ability to cross-learn, or re-establish coherence between the two systems, to perform deductive and inductive reasoning, and to build a reification process that applies to all these processes and will allow AI systems to reason on reasoning.

Four Essentials

The ability to explain:

It is quite understandable that a System1 would not be able to fully explain itself. After all, we can't always explain our "hunches" about people, situations and decisions. However, a System2 must have this property, as it is the way we communicate with others about why we decide what we think or do. As humans, we have the ability to attempt at re-establishing coherence between our System1 and System2. We learn early that putting our hand on a hot stove is painful. We learn later how stoves get hotter to enable cooking, and that it is a bad idea to cook our hand with excess heat, so that we can later teach our children why it is a bad idea to touch a hot stove.

This shows that System1 cannot be a pure black box manipulating inexpressible conceptual structures. It needs a bridge to System2, and that bridge uses words. For some coherence to have a chance to exist between System1 and System2, these words must be part of a meaningful structure, called an ontology such as the one we built in the Causality Link platform.

The ability to forget selectively:

Humans often decide that a specific author is not credible as they reassess the validity of the statements that this author made, therefore actively lowering the value of pieces of knowledge that have been deemed unworthy. LLMs cannot do that, as the training process blends all content into a single interpretation network. The solution with current LLMs relies on the building of different networks with different corpuses. This makes the evaluation of the quality of the additional information provided by a part of any corpus very difficult. The Causality Link system can dynamically discard part of the corpus it leverages, which enables dynamic comparisons of results achieved with or without parts of the corpus of texts provided to the system. This can be used for example to dynamically remove publishers who incorporate an increasing quantity of repetitive AI-generated content to their feed.

The ability to learn continuously:

A major limitation of current LLMs is that they can only learn through expensive training periods rather than continuously. This drawback was seen early on and led to the invention of the RAG concept to try to account for newer information that arrived since the last training period of the LLM, or to add information that trainers did not have access to. While many researchers have contributed immensely to variants of RAG, these efforts are just trying to fix the original issue: LLMs today do not update their knowledge continuously, like humans do. LLM research must solve that issue, especially in fast-moving domains such as finance or geo-political analysis.

The ability to go back in time:

To understand how good a decision system is, we need to perform "back-testing", that is an analysis of the performance of the system at a given point in time in history, to assess the gain in performance we would have achieved had we had access to such a system in a previously known context. This is not possible with the current LLMs, which respond with the knowledge made available to them during their last training. Of course, it would be possible to re-train them only with documents available on that date in the past, but this, like the ability to forget, entails costly and tedious corpus manipulations. Because the Causality Link symbolic detections are indexed by the date at which the paper that contained them was published, we were able to develop a system with a dynamic "Time Machine" where it is possible to choose any day in the previous 10 years and analyze what recommendations of the system would have been at that time. Interestingly enough, we also found that humans only have one single "future" in mind, which is the current one, and that it is extremely difficult to remember what precise future we envisioned at an arbitrary time in the past. This feature is therefore not necessary to mimic the human brain, but it is necessary to validate the quality of recommendations that such a system would have made in the past, a crucial element to solidify our trust in such a system.

Causality Link's Data Model

The Causality Link vision to achieve these objectives is built on the extraction of readable patterns from texts and their storage in a data lake, rather than by letting the convolution algorithms condense the extracted information into a difficult to decipher neural network. The commercial product offers four such patterns while two more are in the research phase. These patterns can be conceived as mini knowledge graphs built on top of entities with the accepted meaning of entity in "entity recognition".

Entities are of many types for our domain: company, product, industry, location, author, date, trend type, causal link type, etc… The four main patterns that we extract from texts are built on these entities and are called "*indicator*", "*trend*", "*event*" and "*causal link*".

An **indicator** represents the definition of a time series and is built by complementing a key performance indicator (KPI) such as "GDP" or "sales" or "EBITDA" with some other co-referenced entities whose type is statically defined by the KPI and is part of the ontology. For example, "GDP" will require a country in the context to become a fully formed indicator such as "USA-GDP". Indicators have a date at which they are supposed to be true, and a publish date. When the publish date is later than the true date, this is a statement about the past, when the reverse is true, the indicator is a forecast. Indicators can be about very detailed data, such as "sales of Tesla Model 3 in China on Q4 2024", and our model uses the taxonomy of each type of entity to enable different aggregations of indicators, such as "sales of Tesla Model3 in Q4 2024", "revenue of Tesla in China in 2024", "Tesla revenue in 2014", etc… This aggregation capability is essential to nowcast multiple statements into a continuously evolving value for an indicator. The readability of entities and indicators combined with the taxonomies enables full transparency of the generalization/aggregation process.

A **trend** represents the evolution of an indicator over a specific period. It is therefore built on top of an indicator, with one additional date and a delay: the initial date marks the beginning of the measurement period of this indicator, and the delay measures the comparison span. Trends are very used in finance, with quarterly trends being the most common, and the choice to compare two consecutive quarters (in this case, the comparison span is one quarter) or the same quarter year over year (the comparison span is one year). Trends are used to nowcast the evolution of any indicator, using the Positive Trend Percentage (PTP) concept: the percentage of statements that mention a positive evolution of that indicator.

Events represent non-continuous concepts, such as pandemic, war, natural or artificial catastrophes, etc… Just like indicators, the detection of events requires the equivalent of a KPI, called a baseEvent, and the entities that complement it. At the difference of indicators, events can require two entities of the same type. For example, a corporate acquisition requires two companies, a war requires two locations, etc…

Causal links represent the explicit indication in a text that an instance of one of the previous three types impacts another instance of one of these three types. We have so far detected over 60 million such causal statements in our 160 million texts, ranging from "The Russian war in Ukraine impacted negatively the production of electric harness for Volkswagen", to "Any 5 % increase of Ford F150 translates into a 10% growth of Ford's profit in the quarter". In the last example, we are able to extract the elasticity of the indicator "Ford-profit" against the indicator "Ford-F150-sales". Aggregating these causal links enables the creation of causal models as large graphs linking hundreds of relevant indicators and events together, at the micro or macro level.

To extract these patterns from texts, we use an unconventional NLU pipeline, which leverages a large proprietary ontology describing more than 3,000 KPI, more than 1,000 types of events, most locations in the world, over 40,000 companies operating in over 3,000 industries and listed on more than 200 public exchanges around the world. This ontology is evolving rapidly by leveraging multiple proprietary tools as well as LLMs, and is a parameter of the NLU process, meaning that every text is analyzed by first loading the current version of the ontology. This separation between the NLU code and the semantics of the domain encoded in the ontology enables both a rapid evolution of said semantics, and a compaction of the NLU code which is based on a black board architecture of meta-rules that are not aware of the semantics of the domain on which they apply. These rules are contained in two sequential processes, a synonym process and a pattern-matching process. The synonym process takes advantage of the synonyms defined in the ontology, with complex "contextual synonyms" enabling the disambiguation of most entities. The pattern-matching process uses a few thousand Prolog rules that build incrementally the desired patterns over the words of the analyzed text.

Extracted patterns are stored in a hot NoSQL database and also exported hourly into partitioned S3 buckets using a compact AVRO format. The resulting data lake can then be queried using SQL powered engines including Presto and Spark; accessible via Jupyter notebooks for example or accessed transparently through a SaaS product which encompasses hundreds of predefined queries.

Real Time Learning in Action

Let us scan how this architecture contributes to the four features required for future LLMs.

First of all, because all aggregation results are built on top of patterns which point back to the sentence and text from which they are extracted, we can easily produce the quotes corresponding to these patterns, and the texts from which these quotes were extracted. This enables for example a Structure RAG architecture where analytics on patterns are used to select the right quotes/texts to provide to the LLMs for summarization, rather than using distance computed with vector databases: https://causalitylink.com/resources_/analytics-controlled-narratives-with-srag/ .

Then, because all results are dynamically computed from the data lake using patterns that point back to the original texts, it is quite simple to refine these queries in order to remove some specific parts of our corpus, and for example compare results achieved with or without these specific documents or publishers. Delaying some of the processing to query time enables very easy manipulation of the corpus content, including forgetting some of it.

It is also clear that such architecture is learning continuously, as texts are analyzed in real-time, patterns are then stored in the data lake and become available in real time, or near real time, depending on the specific needs of the product.

Finally, because all our detections contain copious timestamps for the original documents, and because queries perform the aggregation at run time, it is quite easy to filter detections up to a certain date, and provide the precise answer that the system would have provided at that date if that query had been executed, which enables a back-testing ability that is very important for the validation of the global model.

Sample Neuro-symbolic Results

Because we had a significant symbolic platform at our disposal when the LLMs arrived, we were able to be among the first to explore the benefits of combining features of both architectures. Here are a few examples of what has been achieved.

Improving our ontology

There are a very large number of ways to improve an ontology with a LLM, or more precisely, to confront the coherence of both tools for their mutual benefit.

A very easy step consists in very <u>local improvements of the ontology's taxonomy</u>, by querying a LLM for the list of the children of each concept at every level in the taxonomy. Comparing the taxonomy list of children, and the results of the LLM leverages the LLM for completion of the taxonomy, or more simply to discover new synonyms for existing children. This generated a significant acceleration of the speed of development of our taxonomy with much less manual work. We also discovered bizarre errors in the

LLMs, such as for example that the vector distance between “revenue” and “EBITDA” was smaller than the distance between “EBITDA” and “profit”, for some version of GPT4. Such an error could lead to wrong aggregation by the LLM of the signal available for “profit” and “revenue” for example.

More recently, we leveraged the pattern detection ability of a LLM to compare the results of our “trends” detections and the results of a LLM when provided with the same sentence from which we detected our trends. This enabled us to complement our synonyms for a fraction of our taxonomy of trends, and to detect faster a few sources of misunderstanding for the LLM we used.

Improving on RAG to create SRAG:

We have described in the paper mentioned above how to filter documents for a LLM to summarize using our symbolic architecture which replaces a traditional RAG before passing those filtered documents for summarization to a LLM in order to achieve more coherent and less hallucinatory results.

Generation of real-time alerts:

Combining real-time learning and a symbolic description of the patterns learned, it is possible to perform an early detection of new topics trending with simple data analytics on these patterns. This has enabled us to produce an alert detection mechanism which compares the frequency of any mentioned pattern on a given day with that of the previous 100 days. We found that this detection of new causal relationships would have given an advanced warning to companies leveraging it at the beginning of the 2020 Covid crisis as well as when at the onset of the 2021 semi-conductor shortage: https://causalitylink.com/resources_/beyond-the-noise-the-power-of-pertinent-market-alerts/

Contrarian analysis agents:

We have described in another paper https://causalitylink.com/resources_/contrarian-analysis-agent/ the construction of a “contrarian analysis agent” that plunges queries on our symbolic data lake into the API of an OpenAI agent, in order to build a system that could filter, from all texts, the ones mentioning KPIs of a selected company or country that were the object of a consensus on their future trends, and had also a small number of contrarian viewpoints. A LLM then summarizes these results, with a focus on contrarian viewpoints, enabling a rapid filter on early contrarian signal detection.

Automatic generation of Bayesian Networks:

Using the causal graphs extracted from millions of texts, we have developed the ability to automatically transform these graphs into Bayesian Networks and leverage their predictive capabilities in finance.

The results were published in the Journal of Financial Data Science in April 2022 https://www.researchgate.net/publication/359967807_Building_Probabilistic_Causal_Models_Using_Collective_Intelligence , and this paper achieved third prize in the 2023 ADIA “Causal Research in Investments” competition https://www.adialab.ae/call-for-papers .

We then applied successfully this method to predict the future evolution of the price of crude oil in the 2020-2023 period, demonstrating that there is value in the collective intelligence of journalists worldwide. A paper covering these exciting results is available by request.

Conclusion

We are at a pivotal moment where modern machine learning research can benefit from the previous generation of symbolic AI and combine the advantages of System1 and System2 into true neuro-symbolic architectures, using an agent framework. Once LLMs achieve the ability to explain, forget, learn continuously and go back in time, the intellectual power of these neuro-symbolic architectures will become truly super-human.