

IRIS.AI

Plausibility and factuality of LLMs

Tech angle

CuttingEdgeAI: Large Language Models

● founders@iris.ai

● www.iris.ai

Current LLMs focus: **plausibility, coherence and paraphrasing**

When successful the generated text reads well and looks similar to human's.

And **NOT** focus on a concrete fact as a target

VI

Is vanadium a good conductor of heat?



Yes, vanadium is a good conductor of heat, although it is not the best among all metals. As a metal, vanadium has a crystalline structure that allows the movement of free electrons, which contributes to its ability to conduct both heat and electricity.

- **Paraphrasing**
is great but not for phone numbers, or years, or material properties
- **Most probable word prediction**
works well, but not for endemic words
- **Regular language tokens**
are used comprehensively, but numbers and entities' context have different meaning

Leads to
factual correctness
issues



Hallucinations

Phenomenon when the machine invents statements, claims or facts that are not real.



Mixed contexts

Mixed contexts is usually when a contextual factual information is wrongly linked to a subject that is similar but not the same as the original.



Biases

Biases have many forms and are usually caused by imbalanced datasets, lack of examples or too big of a contextual diversity.
representable

Solving **factuality** makes the plausible truthful



Structured knowledge

LLMs work with unstructured knowledge expressed in language, but a lot of human knowledge is structured and described in structured sources.

How to use it today, and how to leverage it in the future?



Smart Large Data

Currently the understanding is that sheer volume of data is representable for the knowledge to learn.

ChatGPT was successful because of its safeguards and data curation.



Changing focus

Attention in transformers provides great tool for in-context learning.

Next level attention needs to vary depending on the detail required for the predicted token

structured knowledge

is necessary for metrics, validation and short term corrections via post processing

**7 years of
R&D**
**at the forefront of
NLP and machine
vision.**

At Iris.ai use existing structured knowledge

- ✓ To inject it into existing models
- ✓ To create automated metrics for measuring hallucinations and mixed contexts
- ✓ For validation

Creation of structured knowledge

Via humans is expensive and non-scalable
We use our Extract Tool to structure knowledge from scientific articles at scale

is important to fight biases, to balance contextual representations, to facilitate contextual diversity

**7 years of
R&D**
**at the forefront of
NLP and machine
vision.**

Using Iris.ai's ensemble of models

- ✓ We can analyze datasets for misrepresented tokens
- ✓ We can identify tokens with diverse meanings
- ✓ Thus we can provide a dataset curator a viable information for balancing datasets

Our Explore, Extract and Analyze Tools

Could and are used today to automatically create large semi curated datasets with balanced contextual representations

Feeding those datasets to a next generation of models could improve these capabilities as well

IBM Watson AI **XPRIZE**

Top 10 in AI for Good XPRIZE



Powerful Academic research
partners

Being able to
change focus

is vital for getting facts correct on the fly.

**7 years of
R&D**
**at the forefront of
NLP and machine
vision.**

Changing focus requires more R&D

- New metrics need to be developed to allow experimentation with training objectives
- Experimentation with the attention mechanism
- Experimentation with new architectures

Cost should not be underestimated

Structured knowledge, good data and changing focus should improve factuality

But the resulting models should be optimized and viable to modify, fine-tune and re-train

IBM Watson AI XPRIZE

Top 10 in AI for Good XPRIZE



Powerful Academic research
partners

Roadmap to Science

Question-and-Answering machine

- 1 Establishing new metrics for Factuality
- 2 Curating new datasets based on Structured Data
- 3 Knowledge Injection
- 4 Integrated Factual Correctness
- 5 New LLM trained



Plausibility + Factuality

When successful the generated content will revolutionize how we use science and how we use information in general.

VI

Is vanadium a good conductor of heat?



No vanadium is not a good conductor of heat. Vanadium is an average-hard, ductile, steel-blue metal. Unlike most metals, it is electrically conductive and thermally insulating.

Focus on all

- Plausibility and Paraphrasing
- Factual correctness
- Limited biases and safeguards

Future work

- Know your audience
- Change explanations by figuring the knowledge and vocabulary level of the user

IRIS.AI

**Unlock a new era of
scientific discoveries
in collaboration with AI**

● founders@iris.ai

● www.iris.ai