

Why we don't have Explainable AI

Rajagopal Tampi

It has been more than three years since the launch of Chat GPT set the technical world on fire. Since then, I have written many articles on the urgent and important need for explainable and human-centric AI. The subject is covered in my book¹. The world has been grappling with hallucinations, deception, Machiavellian behaviour and other such harmful tendencies of AI, yet nor have the largest MNC vendors of LLMs or Governments of nations taken steps to ensure the building explainable AI.

Recently it captured my full attention when I read about a mathematical reason for these errors in AI. The author is a mathematician and the article² deals with purely mathematical reasons for the unexplained errors of AI. The author's arguments are logical and make sense. The mathematical formulae and processes suggested will help rid many of the AI errors we face today.

We have algorithmic coders but they are not all mathematicians. Nor do coders think in broad terms linking formulae used with possible effects in unconnected domains or this problem would have been highlighted well before today. They focus on the straight and narrow path of coding, providing the working code to perform a mathematical operation.

The right way to eliminate all doubt about fake minimums, saddle, troughs and directions of change including the rotation around the three axes is to use the Hessian (second derivative matrix) of the loss function. The eigenvalues condition number (κ), the eigenvalue magnitude (ϵ) and the negative eigenvalue count (δ) will accurately confirm the changes in the Hessian.

None of the LLM vendors use Hessian. The reason is that the iterations in Hessian are a 100 times more than the commonly used gradient descent algorithm which requires approx $K = 100$ iterations to cut the error in half. This means that using Hessian we will be burning GPU power and compute bills will shoot up dramatically. Imagine using Hessian to optimize a loss function in an LLM (Chat GPT 4) which has 1.8 trillion parameters. The astronomical amounts of compute power required are not practical or even affordable. That's why Pytorch, TensorFlow, Jax and Keras, the tools of MNC AI vendors do not use Hessian.

Surely these companies are aware of the Hessian mathematics discovered in 1950. Yet even if they cannot use it today, they choose to hide the true reason from their customers, governments and people.

Why not slow down the development to build explainable, high quality and reliable AI using currently affordable economics, technology and mathematics with zero errors? Is that not a better path to take? **The LLM vendors led by Chat GPT do not agree.**

To choose to ignore this known reason which is making AI error prone and unexplainable is indicative of a serious ethical malaise in the industry.

With trillions of dollars of investment being poured into AI, the least we expect is honesty about why there is no explainability, why hallucinations exist and what is being done about it. AI

applications are pervading medicine, healthcare and many other life threatening and sensitive fields. People will suffer because of these errors.

Governments have not held the companies accountable. The situation indicates the low levels of quality of the industry, political leadership and societal degeneration in our world.

Reference:

1. [Applied Human-Centric AI](#)
2. [Three numbers. That's all your AI needs to work](#)

Download a pdf

[About the Author](#)

Disclaimer: The opinions expressed in this article are personal opinions and futuristic thoughts of the author. No comment or opinion expressed in this article is made with any intent to discredit, malign, cause damage, loss to or criticize or in any other way disadvantage any person, people, company, government, country or global and regional agencies.