So You Want to Use AI in Engineering Education? Bad Idea!

April 8, 2025

Dr. Kristinn R. Thórisson

Professor of Computer Science at Reykjavik University, Assistant Director of Reykjavik University's AI lab CADIA and Managing Director of the Icelandic Institute for Intelligent Machines

Dr. Kristinn R. Thórisson takes us through the historic context of the current genAI hype and why he thinks contemporary AI such as ChatGPT is bad for education.

Two ideas wrestle in the mind of anyone who is trying to the to the bottom of AI at present: On the one hand is the idea of machines that might be said to be "intelligent," within common-sensical bounds; on the other hand is the application of a variety of ideas that AI research has produced to date to automate a variety of everyday things, graphic design, software programming, text generation, factory processes, or other predefined activity. These two ideas are related, of course, but they are not the same thing. One of them refers to the future – a future that still has not arrived: Machines that think; machines that are intelligent. The other refers to the application of what has been found out so far. The former feeds the latter, not the other way around. Failing to separate them results in confusion and incorrect predictions.

One of them being that contemporary applied AI technology is great for education. Today's chatbots are trained on a substantial amount of digital content – text, of course, but also images, videos and audio recordings. Even in their otherwise excellent role as pattern extractors and pattern generators, artificial neural networks, generative AI, and related technologies are significantly limited in their abilities. ChatGPT is overconfident in its answers and inherently incapable of realizing when it makes

mistakes. It makes approximately mistakes at least 4% of the time – one out of every twenty five answers are wrong. This well-recognized fact has so far gone unexplained, and large numbers of people are working on finding solutions to it. The most commonly proposed one used to be, and still to some extent is, that more data is needed. However, improving accuracy with more data has hit a ceiling. Not only is it extremely expensive (US tech giants are considering opening closed-down nuclear plants to generate the necessary power), we are running out of data. There is another, theoretical reason for this ceiling effect: The systems do not have any mechanism to self-correct when they get things wrong, because the statistical principles on which they manipulate the information they are given do not permit it. To do so requires verifiable knowledge of cause-effect.

Without a deep understanding of cause-effect relations, engineering would be impossible: A professional engineer can ensure that a bridge doesn‘t fall down, and that a skyscraper withstands certain windspeeds, by the logical application of knowlege at many levels of detail, each one resting on well-known principles of cause-effect relationships, given that certain conditions of e.g. manufacturing and weather hold. Ensuring that those conditions hold is based on cause-effect reasoning chains as well – some might call much of that “common sense.“ We humans make use of such reasoning chains every time we make plans, infer what someone means when wink at us, or simply when we try to find what we did with that damn TV remote. Contemporary AI can only represent information statistically, they contain no functional representation of cause-effect relation. As a result, it is quite a stretch to say that they “reason“ or “think“ – if we dare make this claim it cannot possibly be meant literally.

When an expert tells us they use this technology to save time when writing articles, or to find relevant material online, they are very likely telling the truth – but make no mistake: It‘s not the AI that‘s the expert here, its the human. The technology solution in this case is less like asking an expert for advice and much more like using an online search engine. With sufficient expertise in the field in question, the human user is able to judge the quality of the output of the AI system and choose which suggestions to use, rely on, trust, follow up on, and which to discard. Their confidence clearly outshines the built-in overconfidence of these systems because they are applying them as a tool in their own field of expertise. If the number of errors that these technologies make were extremely low, like one in one million, or even one in one thousand, a potential justification for augementing the work of a teacher with this kind of automation could be that the benefits outweigh the downsides. Students are by definition not experts in the subject they are studying, so contemporary AI should not be used as a replacement for good teachers.

So this is the bad news I bring you: Contemporary applied AI technology is useful for those who already know what they‘re doing – who already have ingested the information that the AI is being used for and can judge whether and when to trust it, use it, pursue it, or discard it.

There is more bad news: Contemporary AI doesn't work very well – or even not at all – when the necessary data for training it is in short supply. Generative AI and other similar technologies require big-data, otherwise they don't work. To take one example, the Icelandic language is a "small language." While the amount of content that has been written in Icelandic is not limited to the Icelandic Sagas, it simply is insufficient to make ChatGPT, which has been deliberately trained on the language, fluent enough in constructing grammatically correct sentences in even short paragraphs of text. This is not just text – Iceland is a small-data nation: Even the largest holder of data in the country, Statistics Iceland, must admit this. The Icelandic census from 1703 to 1920 holds just under one million records, and it contains at least three records for everyone alive during that period. Trying to use contemporary AI to infer e.g. who lived where when is not possible because the data on individual persons is too scarce. On most counts, the Nordic countries share this very same problem.

Despair not, however! AI will get better – as academic research progresses, better AI methods will emerge. The next 10-15 years will see new kinds of AI architectures that can work with actual causal relations, reason and plan using causal chains whose reliability can be counted on. These systems will be able to truly explain what they are doing, plan to do, and have been doing – and why. Unlike today's applied AI, they will give verifiable arguments for their decisions and we will all agree that they deserve our trust. Such an AI future cannot be ushered in solely through the efforts of private companies, whose horizon seldom reaches beyond 4 years and whose work primarily involves applied AI, not the creation of truly new ones. With proper funding for academic research in artificial intelligence, a future of trustworthy automation could arrive in a decade; without such funding it will be significantly delayed. In the mean time, I hope that we can avoid destroying the education of our youth by an over-reliance on untrustworthy technology.

Dr. Kristinn R. Thórisson is Professor of Computer Science at RU. During his 30+ years of AI research he has worked at MIT, LEGO, and British Telecom, founded several startups, taught AI courses at Columbia University, RU and KTH, and consulted on robotics for NASA and HONDA Research Labs. Dr. Thórisson has been an advisor on AI to the Prime Minister of Iceland and the Swedish government. He is a three-time recipient of the Kurzweil Award for his work on artificial general intelligence.