Exploring AI in Indian school education

Gurumurthy Kasinathan *
Yogesh K S †

20 March 2019

Abstract

Artificial Intelligence (AI) gets its intelligence by learning patterns from large volumes of data. The application of AI in education generally requires data from curriculum resources and relevant metadata to develop useful intelligence. While AI in education in India is still in its nascent stages, it is more advanced in the field of health, which can be compared with education in its essential nature and purpose, as far as the development of AI solutions is concerned. There are several challenges to the use of AI in education. Some are common to the use of AI across sectors. One such challenge is that the algorithms constituting the AI are not neutral or objective. Biases of the programmers get into the code and get amplified in the machine generated solutions. The existing biases and harmful approaches like rote or drill learning, inappropriate curricular materials and privileging tendencies over inclinations/interests will get amplified through AI.

Democratizing AI will help in mitigating these challenges which can save us from a Kafkaesque future. The first issue is the ownership of data collected about teachers and students. Currently in the absence of any frameworks,

*Gurumurthy Kasinathan is Director of IT for Change (`https://itforchange.net`) and leads its education programs.

†Yogesh K S is Project Associate - Technology with IT for Change, and as a FOSS evangelist, leads the technology implementation across projects.

such data may be under the possession and de-facto ownership of the IT company providing the solution. The biggest contribution that AI could make to education in the Indian context would be to promote the idea of learning as learner-centred. This means that in the curricular content, the teaching methods would need to be adapted by the teacher based on the needs and contexts of each learner, moving from the one-size-fits-all approach that is dominant and expanding the limited content and pedagogy practices of most teachers. Any change in education has huge impact - affects future generation and society and hence the use of AI in education should be done only with adequate attention to the issues raised.

Keywords: artificial intelligence, algorithmic bias, machine learning, education in india

1 Background

While the idea of Artificial Intelligence (AI) is not new, its emerging avatar, based on the three underlying phenomena of big data, machine learning and exponential increase in computing power has enabled AI to become a reality. Algorithms created through machine learning are being able to meet the Turing tests, of exhibiting intelligent behavior indistinguishable from that of a human, covering high-level cognitive processes like thinking, perceiving, learning, problem solving / decision-making etc.

With advances in data collection and aggregation, analytics and computer processing power, Deepmind, the AI engine of Google defeated the world champion of the game Go, a very complex game, one of several instances to show that AI can even be equal or even superior to human intelligence.[1]

The Niti Aayog, Government of India, has published a Discussion Paper on *National Strategy for Artificial Intelligence*, in which it has identified the focus as "AI for All" to leverage the transformative technologies to ensure social and inclusive growth in line with the development philosophy of the government. The paper has highlighted Education as one of the core areas of AIs focus in India (others being Healthcare, Agriculture, Smart Cities and Infrastructure and Smart Mobility and Transportation). The paper aims to explore potential of AI for impacting educational processes and outcomes

(both positive and negative).[2]

2 Context of Indian education

Indian school education system is among the largest in the world. The public education (government) system in India is however weak and ill resourced. It is battling old challenges of inadequate infrastructure, poorly qualified and motivated teacher force (with teachers facing isolation and lack of trust), as well as newer challenges of increasing privatization and commercialization of sector leading to stratification. The diffusion of digital technologies has also been uneven, adding to the stratification.

A pedagogical challenge that is an important cause of low learning levels in Indian schools (both government and private), is the reliance on rote-memorisation of content, as opposed to meaning making, though the latter has been emphasized in policy and curricular documents including the National Curricular Framework 2005. Teachers see their aim as completing the syllabus by ensuring text book contents are memorized by teachers and reproduced in written examinations, which are the primary mode of assessing learning. In this process, many students do not understand the concepts that the content is supposed to communicate. ASER studies point out that a large percentage of students in grade 8 are unable to read a grade 2 text or solve a grade 2 arithmetic problem.

Secondly due to large class size and inadequate number of teachers corresponding to number of grades and sections in a school, teachers are unable (and/or not motivated) to connect to individual learners, understand their learning levels, learning needs and interests, to connect students to the learning process. Content, pedagogy and assessment follow the one-size-fits-all approach, leading to poor learning levels.

3 AI in education

The starting point for AI in its current avatar is the availability of large volumes of data, the code of which is written by humans, and is further generated by machines which mine and analyze the data to derive patterns

which can be refined by running these through even more data. In education, such big data would comprise educational content developed by teachers, educators, education systems and experts, banks of teaching approaches / methodologies, assessment banks; these data are relatively static'. Dynamic big data includes information on transactions (lessons and activities) and student responses to activities and assessments. This big data can be used to develop algorithms that can support:

1. Self-learning through adaptive practice

2. Teaching through personalized education where custom content, pedagogy and assessment can be derived for each student based on her/his responses to past activities and assessments

3. Macro diagnostics and predictive models, across groups of learners (by geography, demographic profile, grade, medium of instruction, subject and other categories)

The aim of an entity designing an AI application in education would include the task of capturing large volumes of curricular resources, with metadata relating to the subject area, relevant grades, language, type of resource, format of file, level of resource, role (introducing a topic, reinforcing learning) in teaching etc. NCERT for instance has listed a set of 31 such metadata elements to be tagged to each resource available in its NROER (National Repository of Open Educational Resources) repository.

Another task would be to capture actual use of these resources in different learning situations, the activities of the teacher (pedagogies) and the responses of the learners to the content, usually through assessment process that seek to ascertain the level of learning/understanding of the student before and after the transaction with a resource.

Analyzing if the learner has learnt / understood the concept, the feedback for that particular resource unit and the transaction approach would be recorded. Recording thousands, millions of such responses to different resource units (along with the profile of the teacher and the learner) for a concept and combinations of these resource units with different transaction methods, with the feedback (on its effectiveness in ensuring learning) would help the algorithm identify the effective learning resources and methods given

a learner context. Identifying the few correct content along with transaction approaches would improve teaching efficiencies as well as effectiveness, is the assumption.

The same can also support self learning. A student can use resources provided by the AI engine and respond to the assessment activities, providing feedback on the learning from the content and activities provided. Based on the analyses of the feedback, the AI engine would suggest, provide further content and activities to the learner.

4 AI in education in India

While AI in education is still in nascent stages, it is more advanced in the field of health, which can be compared with education in its essential nature and purpose, as far as the development of AI solutions is concerned. Health aims to identify issues/problems with the human body, and based on past data collected of similar problems as well as solutions provided (and responses to these solution), keep refining its repository of solutions for each of these problems (with as much data, metadata about each problem). Over time, as more and more data is collected, the number of variations in health problems and possible contextual remedies will be captured and the AI will be able to provide a most relevant solution for a given problem. For instance Sloan Kettering Memorial Hospital is attempting to build such AI using its hospital data with the Watson Oncology Advisor.[4]

Similarly, in education, the problem of curing ignorance (promoting learning), content and pedagogies to address this ignorance and assessment data providing evidence of success/effectiveness of these content/pedagogies can be collected. Over time as more data is available on the different challenges (learning requirements, misconceptions), and solutions (custom content and pedagogies), the AI will be able to suggest the appropriate content/pedagogy for any learning requirement, assessing the learner context, profile and aptitudes.

In the USA, the approaches outlined above are already being used by AI applications such as Watson Teacher Advisor (from IBM) and G-suite for Education (from Google). The widespread use of personal digital devices

by students provides vendors like Google to collect large volumes of data. While full fledged personal analytics and personalized learning is still some time away, the approach of the current AI models gives us a basis to imagine a future where the AI algorithms would understand the learning levels, misconceptions, learning styles, interest areas, needs/priorities of each student and based on these provide custom content, pedagogies and assessments for each learner.

In India, since digital education in schools is still long way off from being universalised, thanks to basic infrastructure, maintenance, electricity issues, data collection is still in its nascent stages. A company that has developed a product called Mindspark claims to have data from assessments which run into millions of tests for lakhs of students.[6] If a student is able to complete a two digit addition sum (without carry over), he could move to the next step of addition with carry over. If the student gets the two digit addition wrong, then more two digit sums and /or single digit addition sums can be provided to assess if student has understood the concept of addition. Text problems more connected to the life of the student can be provided instead of plain numeric sums. This is provided as an example to help understand the working of the AI engine. Thus, AI can help in identifying relevant learning methodologies for diverse contexts through use of big data, thereby enabling improved quality of education.

Government of India collects micro level data on students every year which is available on the UDISE (Universal District Information for School Education). However the government has not yet developed any AI infrastructure for analyzing this huge data. This information is collected only once a year, and is therefore static but could be analyzed to derive directions for provision of school infrastructure, teachers as well as possibly, predictions of potential drop-outs based on demographic profile (socio-economic factors) information.

5 AI in education challenges

There are several challenges to the use of AI in education. Some are common to the use of AI across sectors. The first one is that the algorithms constituting the AI are not neutral or objective (assuming there could be

such a thing). Biases of the programmers get into the code and get amplified in the machine generated code. This sometimes becomes clearly evident in the outcomes, such as the cases of criminal profiling algorithms being more (and unfairly) severe on African-Americans in the USA, as also areas such as credit scoring, insurance premium determination etc.

In the Indian education scenario, this risk is very severe. The predominant folk-pedagogies in India privilege education in English medium, over the local/regional language, focus on cracking the examinations than in understanding concepts/making meaning, this is an important cause for rote memorization as the predominant mode of learning. AI is also likely to promote such beliefs and further impoverish learning possibilities of students, especially those belonging to marginalized sections.

Secondly, the current syllabi, in many cases reflects the urban, upper caste, upper class, Hindu and male bias of the syllabi developers. For instance, though agriculture is the livelihood of a majority of the population of India, agricultural practices and ideas have very little place in the text book. Derivations of such content would also reflect these biases. The teacher population profile also reflects these bias, with teachers in government and aided schools, increasingly belonging to a class and caste different from their students.

On the societal side, India is a rigid and feudal society, with widespread problems such as marginalization on grounds of caste, gender, religion and class. The caste system in a sense already uses some kind of intelligence to predict that the best chances for the son of a potter would be to become a potter and of a trader, to become a trader. Secondly the people who would work in the AI algorithms design would be consist of more tech biased people technologists or tech-enamoured educators, who are less likely to be aware of the complexities of education.

5.1 Purpose of education

The primary aim of education is to produce / strengthen a sense of agency in the learner. Hence, life choices based on capacities and aptitudes may be inappropriate for a learner and may end up strengthening legacy evils such as caste system. Thus existing biases and harmful approaches will get am-

plified through AI - rote / drill learning, inappropriate curricular materials, privileging tendencies over inclinations/interests.

At a macro level, the purpose of education is to support the learner to become a 'responsible citizen'. This may require the educational processes to question dominant perspectives and beliefs. 'Going against the grain' may often be essential for progress/break through in society, no society can move forward by looking at rear view mirror.

Algorithms can carry the biases of the people developing them. Another challenge is transparency of algorithm. To reduce if not avoid problems of bias, it will be required to subject the algorithms to audit. For this the algorithms should be released as open source for others to review and comment. Google for instance has a project that releases AI algorithms as open source. However, release of algorithms as open source can also increase chances of misuse of such algorithms. Another option would be for specialized agencies auditing the code, on behalf of regulators.

The dominant model of implementing digital technologies has been through proprietary models. These models distort competition possibilities and provide for huge rent seeking possibilities. Thus Microsoft has been the leader in the office automation space, which enabled super normal profits for the company. This has been amplified in the AI space, with most AI development with very few corporations, all American companies. Google, Facebook. Amazon, Microsoft etc.

Traditional arguments for benefits of free markets have emphasized the role of competition in ensuring benefits for consumers. However, due to the network effects we have seen the opposite phenomena of monopolies of oligopolies in the digital space. Google holds a large part of the search share while Facebook is a near monopoly in the social media space. (Network effect is the phenomena by which the more users our product, the more users you will attract. Since Facebook is the most popular social media platform, it is the platform that new users will join. This is because users of competing products cannot connect to users of Facebook due to lack of inter-operablity. This is also a reason why Microsoft office dominated the office automation world with its proprietary software producing outputs with proprietary formats that were not easy to read and edit in other office products, locking in

users into its software.

Indeed, todays market place is so distorted that new entrants try to buy up the market with huge price discounts, freebies, burning up venture capital cash. The venture capitalists fund a company which is making huge losses in its business only because of the expectation of super normal profits, once the company kills competition with predatory pricing and becomes a (near) monopoly. Even in the area of AI, the company that manages to collect and control large volumes of data will be in a position to command the market.

5.2 Mitigation of challenges

To save us from a Kafkaesque future, we must democratize AI[8]. The first issue is the ownership of data collected about teachers and students. Currently in the absence of any frameworks, such data may be under the possession and de-facto ownership of the IT company providing the solution. For instance users of google suite for education in Vishakapatnam corporation schools, would be providing data about themselves as well as their transaction and assessment information would be collected by Google. Availability of this data for the school itself, the teachers and students (both micro individual data, as well as data aggregated for the school) is unknown, as well as to other organizations interested in analyzing the data to refine their AI engines. Frameworks to ensure an ethical AI are essential, these should stipulate data ownership to lie with the school and not allow for monopoly control of data by the company providing services to the school. Second issue is the transparency of the code that analyses the data. The code will hide the biases of the coders and hence it is essential that it should be visible to audit.

The Google search engine is designed to throw advertisements and page rank can be influenced by its commercial interests, the code may consciously or unconsconsiously reflect the commercial interests of the company and the only way to avoid this problem is to make the code subject to audit (volkswagon car diesel exhaust case). Either the code should be released as open source under copyleft licensing (such as GPL) or be provided to specialised agencies for audit (both white box and black box). Only audited versions should be provided to production cases. Beta code not subject to audit should not be provided to schools.

However, specially in the domain of education, there is a strong case for code transparency as students and teachers can also explore and study the code to learn about AI. AI should not be a blackbox for a place of learning. Teachers and students can not be treated as passive end users of a technology they have no idea how it works. While, there are challenges to understand how dynamically generated code (machine learning) can be audited there are advances in this area.

Such audit is essential to identify (and attempt to address) both biases arising out of commercial considerations of vendors and curricular biases of educators. The latter will be contested terrain, and the least requirement would be of transparency.

6 Conclusions

The biggest contribution that AI could make to education in the Indian context would be to promote the idea of learning as learner-centred, meaning the curricular content, teaching methods would need to be adapted by the teacher to the needs and contexts of each learner, moving from the one-size-fits-all approach that is dominant and expanding the limited content and pedagogy practices of most teachers. AI could do this by suggesting diverse content and pedagogy possibilities to teachers.

However, an enabling policy context is essential so that teachers are empowered to understand the various options with respect to content and pedagogy available and decide what to use and how to adapt these before use. Else, AI would become another force dis-empowering the teacher and forcing her to blindly (without any sense of agency or autonomy) accept content and pedagogy choices in the teaching-learning processes.

Any change in education has huge impact - affects future generation and society and hence the use of AI in education should be done only with adequate attention to the issues raised.

References

[1] AlphaGo versus Lee Sedol - https://en.wikipedia.org/wiki/AlphaGo_versus_Lee_Sedol

[2] Niti Aayog Discussion Paper for a"National Strategy for Artificial Intelligence" - https://niti.gov.in/writereaddata/files/document_publication/NationalStrategy-for-AI-Discussion-Paper.pdf

[3] National Curriculum Framework (NCF 2005) - https://ncert.nic.in/pdf/nc-framework/nf2005-english.pdf

[4] S. P. Somashekhar; M.-J. Sepu lveda. "Watson for Oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board" - https://www.annalsofoncology.org/article/S0923-7534(19)35072-0/fulltext

[5] Smith, Matthew; Neupane, Sujaya. "Artificial intelligence and human development : toward a research agenda" - https://idl-bnc-idrc.dspacedirect.org/handle/10625/56949

[6] Karthik Muralidharan; Abhijeet Singh; Alejandro J. Ganimian. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India" - https://econweb.ucsd.edu/~kamurali/papers/Working%20Papers/Disrupting%20Education%20(Current%20WP).pdf

[7] Draft ethics guidelines for trustworthy AI (EU) - https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai

[8] Stephen Cave. To save us from a Kafkaesque future, we must democratise AI - https://www.theguardian.com/commentisfree/2019/jan/04/future-democratise-AI-artificial-intelligence-power