Cognii
Artificial Intelligence for Education and Training

Dee Kanejiya
Founder and CEO, Cognii
San Francisco, CA

(orig. July 7, 2023)
(edited February 12, 2024)

The Honorable Dr. Arati Prabhakar
Director, White House Office of Science and Technology Policy (OSTP)
Executive Office of the President
1650 Pennsylvania Avenue
Washington, D.C. 20504

Dear Dr. Prabhakar :

I appreciate that OSTP and its National AI Initiative Office (NAIIO) have requested public information about artificial intelligence technology to inform the development of a **National AI Strategy** (RFI DOCKET : OSTP-TECH-2023-0007). As a founder of an AI startup focused on improving the quality and affordability of education, and with a PhD research in Electrical Engineering specializing in language model design and development from the prestigious Indian Institute of Technology Delhi, and with postdoctoral research in language modeling for multimodal interfaces at Carnegie Mellon University and Karlsruhe Institute of Technology in Germany, and having two decades of industry experience in AI, EdTech and natural language processing technology, and as a pioneer of the Conversational EdTech and Virtual Learning Assistant technologies, and as a recipient of an innovation research grant from the National Science Foundation, I am pleased to share my experiential and future perspectives on how AI can play a significant role in the advancement of human civilization, and what specific measures the U.S. government can take to achieve the successful outcomes.

Today AI technology has a potential to transform a large number of industries such as the education, healthcare, justice, and innovation economy. Given my expertise in the technology and education fields, I will focus on the approaches to developing the AI and language modeling technology (RFI Q. 1-2) and its applications to the education and workforce industry (RFI Q. 9-13, 17-23). I will specially emphasize the need and unique opportunity created by AI in transforming the century-old educational assessment practices because everything in a way depends on how we measure human potential. I will also encourage more involvement of the Small Business Administration (SBA) in the AI economy to diversify and strengthen the technological capabilities as well as to democratize the value creation opportunities.

1. AI and Language Model Technology (RFI Q. 1-2)

Artificial Intelligence has been an active area of research since the 1950s with the goal of mimicking human intelligence exhibited across different cognitive faculties such as language and vision. **A language model (LM)** is a mathematical model for primarily assigning a probability distribution over a sequence of words or symbols.

LM Evolution:

1980s : Language models grew popular as a higher order structure to improve the performance of large vocabulary automatic speech recognition and machine translation systems. Early language models were either symbolic such as the rule-based context-free grammars, or statistical such as the **n-grams** which had a narrow context of two or three words.

2000s : Vector representations of words (aka embeddings) started the transition to more complex mathematical modeling of language where each word or its part ('token') is represented by a 100 to 1,000 dimensional numerical vector. My PhD research[1] developed the first **context-dependent vector representation of words** where each word is associated with a different vector depending on its context, using a tensor based model called latent syntactic-semantic analysis, and applied it to build more predictive language models and educational assessment technology.

2020s : The currently popular **large language models (LLM)** rely on more rigorous context-dependent vector representation of words obtained by an **attention mechanism** in an artificial neural network containing many layers. They can be trained in a generative manner to predict a text sequence or an image, and hence are sometimes called Generative Pre-trained Transformer(GPT) models. These LLMs have become feasible now due to the availability of large amount of training data and computational capability.

Computing Hardware Demand:

Graphics Processing Units (GPU) were originally designed for processing images, which are represented by a matrix of data, requiring vector and matrix computations in parallel. Vector representation of words in LLMs has therefore found a great utility of GPU hardware in language modeling, and as the models become larger, the demand for GPUs has increased significantly in the AI industry. In this regard, the federal government's **CHIPS Act** could play an important role in facilitating the AI innovation in future.

Performance Measures:

Language model performance can be evaluated using an intrinsic measure called **perplexity** which is a statistical indicator of how well the LM can predict an unseen text. Its value ranges between 1 and the vocabulary size which can be 10,000 or more (lower the perplexity the better). We had demonstrated[2] in

[1] https://scholar.google.com/citations?user=KeFRrkMAAAAJ

[2] Kanejiya et al. (2004) "Statistical Language Modeling with Performance Benchmarks using Various Levels of Syntactic-Semantic Information." In *Proc. 20th Int. Conf. on Computational Linguistics (COLING), p. 1161–1167. Geneva, Switzerland.* https://aclanthology.org/C04-1167.pdf

2004 that a highly performant language model with 20,000 words vocabulary can be built to achieve a very low perplexity value of **36.37**. It was validated[3] in 2019 when OpenAI's GPT model achieved a perplexity of **35.76** on the same dataset. As LMs become more powerful, the perplexity numbers have decreased even further, and more task-specific measures such as the accuracy of answering questions or task completion rates have also become popular. **Future measures should include model reliability per energy consumption** i.e. efficiency or quality of output per cost of development and operation.

Modeling Approaches:

There are mainly three approaches to LM development — **deterministic**, **statistical**, and **neural** — and they have their own benefits and limitations when measured across different performance characteristics e.g. reliability, controllability, corrigibility, generalizability, explainability, transparency, training data size, computational resources needed, training time, energy consumption, operational costs, task specific validity and optimization, bias and harmfulness, trust, safety, ethical considerations, data privacy and security, intellectual property attribution etc. **When developing an AI policy, it would be recommended that a wide variety of approaches to LM development are pursued** so we can benefit from having the ability to select and deploy a more optimal single or a hybrid LM solution for our unique needs.

Generative vs Analytical AI:

There are two types of AI and LM technologies — **generative** or **analytical**. Information analysis is considered more important than information generation, because analysis is the primary driver of decision making process in businesses and society. Even though we as human beings perform both the tasks of language generation and analysis, many learned people give the advice that the listening (analytical) skill is more important than the speaking (generative) skill. As a result, if we wish to develop AI in our image, **it would be recommended that we do not stop at the generative AI stage, but also focus on the analytical AI.** Generative AI limits a user agency during content construction and makes them more passive content consumers, while an analytical AI offers more user agency and can lead to more complex conversational interfaces which are better aligned to real-life knowledge exchanges between humans. The contrast between the generative and analytical AI can be similar to the example of learning by reading *vs* learning by doing. A generative AI might also be perceived as having an ordaining characteristic especially when it provides inaccurate information assertively, and is also causing some fear of taking away jobs of various types of content creators. On the other hand, an analytical AI can be seen more like a service tool that we invoke when we need to ease our life and increase productivity while still being in complete control. Sometimes a generative model may provide insights for the development of an analytical model, however most of the time, an analytical model is built in a separate way using different modeling techniques than a large LM architecture because there is not sufficient amount of training data available of the analysis process (which requires human decision labeling). A perfect AI companion should have **a balanced high quality generative and analytical** capabilities.

---

[3] https://paperswithcode.com/sota/language-modelling-on-penn-treebank-word

Future Directions:

In addition to improving the performance of uni-modal language models, future advancement of AI should also focus on **multi-sensorial LM technology**, which will simultaneously process multiple sensory modalities for generative and analytical applications. Such systems would combine for example the language, vision, and other sensory information streams for high fidelity human-machine interfaces. We should also consider developing industry or organization specific LM technology that captures the unique characteristics of its domain for reliable analytical applications. Government should also encourage and support the **open-source LLM** initiatives to benefit the small and medium sized enterprises, and develop a larger innovation community.

2. AI in Education (RFI Q. 9-13, 17-23)

Conversational AI offers a unique opportunity to solve a century old problem of assessment and tutoring to make high quality education affordable and scalable. Before discussing the solution, let us understand the background and the problem.

Since the dawn of civilization up until now, natural language conversation has been the primary modality of knowledge exchange between humans. During the ancient times, spoken language based conversation was the only modality for education. Later when writing was invented, spoken and written form of conversation became the modality of education. As humanity advanced and the education system was formalized during the industrial revolution era, the classroom assembly-line model (seat time based credential) became common. Even there the modality of learning was the conversation primarily led by an instructor.

However, one of the drawbacks of the industrial era cohort-based education model was that the students lost the agency and had very little opportunity to express themselves other than during tests. Even that opportunity was taken away a century ago, when the test format of multiple choice questions was invented as a temporary mechanism to recruit a large number of soldiers for the World War I. It is no surprise that the **test format which was conceived and adopted during the time when women did not even have the voting rights in the United States** and racial discrimination was rampant, has been shown to contribute to gender and racial achievement gaps in a number of academic studies[4]. It is also relatively easy to understand the practical irrelevance of the multiple choice tests to our real life e.g. we can ask ourselves : how many multiple choice questions (with precisely four choices and only one correct answer) for performance evaluation we encountered in our regular daily life — today or this week or this month or this year or since we left the education system as a student? The answer is almost zero. Additionally, each multiple choice question with four choices carries a minimum measurement error rate of 25% at the input

[4] Reardon, S. F., Kalogrides, D., Fahle, E. M., Podolsky, A., & Zárate, R. C. (2018). The Relationship Between Test Item Format and Gender Achievement Gaps on Math and ELA Tests in Fourth and Eighth Grades. Educational Researcher, 47(5), 284–294. https://doi.org/10.3102/0013189X18762105

step of a complex assessment process which would likely translate to even higher error rate at the output. These problems have been recognized by all the stakeholders of the education system including the students, teachers, parents, school leaders, employers, and government officials.

Even **President Barack Obama stated during one of his State of the Union Addresses**[5] that *"we need new ways to measure how well our kids think, not how well they can fill in a bubble on a test."* So, it is only fair to ask why such an outdated practice is still continuing to be used after a century and is also consuming a significantly large amount of tax dollars and human productivity? The following questions may lead to a solution: How many different educational testing companies are there with decade-long contracts worth billions of dollars from various state or local education agencies? How many innovative educational assessment startups are being actively solicited by state education agencies? Even the significant amount of funding allocated by the federal government towards Innovative Assessment Demonstration Authority has produced no substantive innovation due to the limitations of organizations involved in its implementation. There are a number of examples of projects funded by the U.S. Education Department where the outcome of more than a million dollars of funding is essentially development of less than ten multiple choice questions and their testing with a few hundred students. The costs of more than $100 million per year associated with the NAEP (National Assessment of Educational Proficiency) platform and assessment development, administration, and scoring could see a significant reduction while increasing its quality if innovative startups are involved.

Why is educational testing important for the civilizational advancement?

The well-known management guru Peter Drucker is often attributed the quote "If you can't measure it, you can't improve it." A corollary to that would be that the extent of improvement would be proportional to the quality of the measure. It is possible that with a poor quality measure, a system might even worsen. During the agrarian era, a foot as a measure of minimum length was probably valid, but if we had remained fixated on it and not developed inches or millimeters, we would not have been able to progress towards the industrial era. Educational testing is a measure of human potential, especially of the developing young human beings who are the future of our civilization. If they are not measured accurately, they will be misplaced in the career education and workforce, resulting in a chaotic society where members do not feel a natural alignment between the work and their skills and interests, and as a result are likely to cancel each other's productivity. On the other hand, if educational testing is perfectly aligned with how human beings are evaluated in real world, which happens through **natural language conversations and work product demonstration instead of asking people to select one of the four choices**, the resulting society will be far better organized with members enhancing each other with complementary contributions. Such a society would certainly be better prepared to tackle the future challenges that humanity may face.

---

[5] https://obamawhitehouse.archives.gov/the-press-office/2014/01/28/president-barack-obamas-state-union-address

Solution to the problems in education:

The good news is that in recent times, the focus has been shifting towards personalized learning and giving students more agency to construct and demonstrate their knowledge. Benjamin Bloom showed in his seminal research on mastery learning[6] that when students receive instruction in a one-to-one tutoring environment and receive instant feedback on their constructed response answers, their performance improves by two standard deviations. This establishes the key role of natural language conversation in the education process and therefore validates **Conversational AI based EdTech as one of the most efficacious educational innovation** available to serve humanity. Cognii has been leading the movement towards Conversational EdTech[7] for the past ten years with its **Virtual Learning Assistant** (VLA) technology supporting schools and higher education institutions in implementing innovative assessment and tutoring solutions. The VLA measures students' learning using high quality constructed response questions and engages them in a tutoring conversation by providing immediate feedback in the zone of proximal development to maximize their learning gains.

My personal involvement with AI in education goes back to twenty years[8] when I presented my doctoral research on intelligent tutoring and assessment system at the very first international gathering of experts building educational applications using natural language processing in 2003 in Edmonton, Canada. Over the years, I have supported this gathering to encourage involvement of young researchers in this intellectually stimulating and satisfying field of human endeavors. A number of other startups and organizations have also played important roles over the years in bringing the benefits of AI technology to the education industry. As someone with expertise in development of both the **language model technology** as well as the **educational technology**, I believe that education industry should not be perceived only as a recipient of the benefits of AI technology, but instead it can take a lead in the design and development of high quality AI technology which in turn could benefit all the other industries. In the recent evaluations of LLM technology, many organizations have started using the human educational tests to measure the intelligence and cognitive performances of AI models. This is both a validation of the importance of educational processes in training an AI model, but also a concern in that the AI models are being evaluated using the same faulty tests that we should be removing from the education system. It is also important to note that currently many of the LLM providers caution against their usage for educational assessment purposes due to their problems with reliability and other limitations. By innovating the educational testing system, we will advance not only the **human learning** but also the **machine learning** field.

---

[6] Bloom, B. S. (1971). Mastery learning. In J. H. Block (Ed.), Mastery learning: Theory and practice (pp. 47–63). New York: Holt, Rinehart and Winston.

[7] https://venturebeat.com/ai/how-ai-will-transform-education-in-2017/

[8] Kanejiya et al. (2003) Automatic evaluation of students' answers using syntactically enhanced LSA, *Proc. HLT-NAACL workshop on Building Educational Applications using NLP*, p. 53-60. https://aclanthology.org/W03-0208.pdf

How can U.S. government facilitate AI driven positive transformation of the education system?
The following policy could provide a possible solution using a two-pronged approach:

i. Every organization (federal, state, local, private) allocating resources for educational testing should require a **10% reduction in number of multiple choice questions every year** or allocate 10% less funding every year for multiple choice tests. This will ensure a graceful transition away from the practically less relevant form of testing and towards more valid and aligned form of AI powered higher quality assessments. On the global stage, it is possible that some developing countries might leapfrog immediately to such a highly efficacious education system with 100% AI powered assessments (similar to their transition to the internet era directly via smartphones and bypassing the computer era of internet) due to the lack of educational testing infrastructure inertia. This will likely create a competitive advantage for them and a challenge for the U.S.

ii. Federal government should leverage the successful practices of its Small Business Administration. Most of the industry sectors have benefited significantly when SBA is involved to facilitate participation of startups to support innovation. Educational testing industry has however remained largely isolated and away from SBA intervention. As a result, there is a stagnation, monopoly, lack of innovation, and sustenance of false or outdated practices which are not aligned with, or could be adversely affecting, the desirable progress of the society and the economy. To address this, federal government should mandate that from every dollar it allocates for educational testing at federal/state/local levels, **at least 20% must be spent to work with startups as defined by SBA**. This will result in a rapid growth of innovative AI powered educational assessment solutions developed by Americans who are bestowed with ingenuity and entrepreneurship. We just need to create a conducive environment for them to demonstrate their innovation. This will increase the technology based economic development, creativity, and competitiveness of the United States. Solving this core assessment problem could also lead to resolution of many of the long standing education problems, as well as the alignment problem of AI technology in general which will lead to responsible and trustworthy AI which in turn will lead to newer and better opportunities for humanity.

I am encouraged by the openness of OSTP and NAIIO in receiving information about AI technology and its various applications as they develop the national priorities and future actions as part of the National AI Strategy. Thank you for this opportunity and I am looking forward to the bright future of humanity powered by AI.

Sincerely,
Dee Kanejiya
Founder and CEO,
Cognii, Inc.
San Francisco, CA