Virtual teaching assistants: A survey of a novel teaching technology

Diane Audras, Alex Zhao, Chuck Isgar, and Yucheng Tang

Abstract
This study centers around virtual teaching assistants (“VTA”) and the potential application of VTAs to reduce the burden on teachers across secondary schools in China exacerbated by the Double Reduction Policy. It seeks to understand how VTAs could be implemented in the classroom while ensuring student and teacher well-being, success, and happiness. Through interviews with fourteen different experts spanning industry, academia, and education, combined with extensive review of the existing literature and ecosystem of VTAs, research shows that VTAs are potentially a major aid in the context of exam review, after school tutoring, automated grading, and student performance reporting. However, they need to be created carefully with an attentive eye towards student support to ensure that the learning experience is not diminished. It must be clearly communicated to all stakeholders that the goal of a VTA is not to replace teachers, but rather to support them such that they can spend more time with students and less time on mundane tasks.

Keywords
artificial intelligence, virtual teaching assistant, double reduction policy, education

Received 8 June 2022; accepted 5 August 2022

Tsinghua University, Beijing, China

Corresponding Authors:
Diane Audras, Schwarzman College, Tsinghua University, Beijing 100084, China.
Email: diane.audras@alumni.schwarzmanscholars.org
dianeaudras1@gmail.com

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
Introduction

The Double Reduction Policy (DRP), known as 双减政策 in China, has sought to ensure equality in education in the country by banning private for-profit after school tutoring. As a result, students need more support while teachers simultaneously have more demands on their time as they seek to provide more after school offerings as required by DRP. VTAs have the potential of helping teachers with their increased workload, while simultaneously creating supportive environments for students. However, their effectiveness should be measured not purely in terms of academic success, but also with regards to students’ mental and emotional well-being.

Burdened by an excessive workload, teachers face their own mental health challenges; VTAs provide an opportunity to lessen both their workload and associated stress levels. Automating the task of grading is a key area where VTAs can play a role in reducing the burden on teachers, especially when teachers lack the time to provide detailed feedback. At the same time, teachers not only struggle to understand the effectiveness of their teaching on the whole, but also to determine which individual students need further help and where this help should be directed. Given the data collection capabilities of VTAs, they could help teachers improve their teaching on a macro and micro level by providing report cards on student performance. This offers the promise of identifying which students need personalized learning support and provide suggestions for that support.

This study will discuss the building of the technology, from the positioning of VTAs in terms of its users and applicability, to its product features. It covers several implementations of VTAs in classrooms in China and broader frameworks for adoption, keeping in mind the limitations associated with the deployment of such novel forms of technology.

At the same time, it is vital that VTAs are not viewed as an end-to-end solution to the burdens placed on teachers. There is a need for education about the VTA itself, with guides that explain that it is not seeking to replace them or their teaching, but rather to be a tool that can hopefully alleviate their burden with tasks like grading. The goal of this is that the teacher can spend more of their time providing individualized support to students. As such, emphasis must be placed on continually gathering feedback from teachers, students, and families to ensure that all parties are receiving the assistance they need.

Shaping technologies, policies and ethics

Education technology and tutoring in China

The education technology sector is huge in Asia, especially the focus on tutoring in preparation for standardized exams (Anderson, 2017). Asian parents are spending considerable amounts of their income on tutoring. Concretely speaking, there are three main categories of online AI-enabled education services in China’s market: teaching assistants, that can assist teachers on tasks such as grading homework and test papers autonomously; teaching management for support on matters such as course scheduling; and assistance for self-study and adaptive learning (Zhan, 2019).

Overall, the literature reveals that tutoring is effective, with some findings indicating that it is more beneficial to have teachers or paraprofessionals (those who might not have been ‘full’ teachers but still have experience in that realm) doing the tutoring than non-expert educators or tutors (Nickow et al., 2020).

Remote tutoring through video calls is one common way of providing education online today. A study in Italy found that tutoring program helped boost tutees’ scores not only with standardized testing, but also increased their “psychological well-being” and “aspirations and socio-emotional skills” (Carlana & La Ferrara, 2021). Additionally, tutees not only improved their scores, but
tutors also scored higher on empathy assessments, showing that there must be a focus on the effects of remote tutoring on the tutors themselves.

Current solutions in tutoring fall into a few different contexts: tutors in small studio environments, complements to large online classes, and hybrid models in between the two. A study attempted to see the effects of establishing a remote design studio for students who were engaging in co-ops or real-world learning experiences (Nespoli et al., 2021), revealing that design studios for students, especially those intended to supplement real-world work experience, are not necessarily a promising opportunity for remote tutoring applications.

In the case of Massive Open Online Courses (MOOCs), there are multiple case studies of VTAs. Jill Watson helps to offload mundane and routine tasks for teachers so that they can engage more deeply with students, while simultaneously helping students focus on what they know and do not know (OnlineEducation, n.d.). Chatbots like Jill Watson are being used to improve the student learning experience and student services in general, providing 24/7 on-demand learning support for them (Teachonline.Ca, 2021). Students are paired with a "lifelong learning companion" that accompanies them to classes and helps with their homework. These virtual friends are not meant to replace the teachers, but instead support them by reducing their workloads (Furness, 2020).

Emotionally speaking, most VTAs currently exist either in the uncanny valley (in which they imperfectly resemble humans, evoking a sense of “uncanniness”) or tend to be more robotic than human. Belpaeme et al. (2017) argue that robots’ appearance and behavior affect learning outcomes, explaining how robots have a distinct social character which influences the student responses that are conducive to learning. Wang et al. (2021) suggest a “mutual theory of mind” as a theoretical framework for designing long-term human-AI interaction, again using Jill Watson as a case study. For instance, the Kruskall-Wallis tests how anthropomorphic the AI is. Automated sentiment analysis can be employed to test the positive effects that the AI’s responses have on the students’ perception of a VTA (Patel et al., 2020).

Technically speaking, robot teachers are not perfect. For instance, in speech recognition and computational vision, there is the risk of inaccurate interpretation of a learner’s behavior (Belpaeme et al., 2018). Moreover, there will often exist gaps between users and developers in expectations and familiarity. Hence, it is crucial for users of chatbots to understand the social and technological limitations of VTAs, to avoid frustration as a result of a mismatch between their expectations and the product’s capabilities (Pang, 2021).

Agile development, a methodology to build software predicated on iterative development and continuous delivery of new features rather than large monolithic releases, is a process for building VTAs. Agile developers can engineer modular microservices communicating with each other to grade homework, provide student feedback, and answer questions. More recently, machine learning algorithms have become a popular approach for VTAs. Reinforcement learning for curriculum planning, natural language processing for grading homework and chatbots, and digital text-to-speech are just a few examples of the types of extant technologies.

**Policy climate in China**

In the summer of 2021, China’s Ministry of Education introduced the Double Reduction Policy in China (Song, 2022). The policy dictates that tutoring classes for subjects taught at school can only be organized on weekdays before 9pm, that tutoring companies cannot use foreign teaching material or employ teachers based abroad, and teach to students below the age of six (Dandan, 2021). The stated goals of the policy are to lessen the financial burden on parents who pay for-profit tutors for services that the state should provide instead, while also helping improve students’ mental health.
China’s existing policy guidelines are focused on strengthening the ability of its research institutions and empowering people to crest the technological wave rather than foregrounding ethical considerations related to AI or education. On the other hand, the country has imposed stringent rules for electronic devices which are used for intelligent education, with some specific policies requiring these products to comply to high ethical standards. There are some tensions between the positive encouragement for AI development and the strict regulations about the implementations and executions of AI, which can also reflect the huge complexity of the ethical and safety issues of AI and education. These in turn might affect the speed at which such VTA technology can be deployed whilst emphasizing the need of adopting a trial-and-error approach given the policy environment.

**Ethical issues surrounding AI in education**

The main ethical concerns for AI education (AIED) are data and computational approaches including data collection, how to analyze the data, whether there are biased assumptions, and transparency issues. The demographic imbalances within AI will influence VTA’s response to the students, because it is trained using posts and data from the previous semester (Eicher et al., 2018). Overlooking these issues concerning minorities may create negative emotional impacts on students, which has guided researchers to develop the system so that it can be as inclusive as possible to a broad range of people.

In-depth investigation of teaching activities in the age of AI has shown that AI teaching assistants may have better cognitive ability, which can relieve the cognition burden for teachers in terms of long-term memory, detail recognition, information storing and so on, but the emotional and creative intelligence of teachers cannot be replaced by an AI teaching assistant system. Despite this, AI teaching assistants are better than teachers in emotion recognition and emotional calculation based on facial expressions (Yang et al., 2019).

Aiken and Epstein (2000) proposed some guidelines when they first set up a series of “dimensions of human being”, including moral, aesthetic, intellectual, and other dimensions, and then argued that AIED should not weaken these human dimensions of students, and at least should be able to enhance one of them.

When creating and launching a virtual teaching assistant, there are ethical responsibilities for those people who are involved in the experiment, and in this case, the responsibility is to deal with tension between authenticity and deception (Eicher et al., 2018). When the researchers used an AI teaching assistant, they didn’t tell students that it was not a human answering their questions, but they justified this decision with three reasons: they didn’t want to influence the students’ interaction with their teachers, students were able to discuss the study at the end of the semester, and those students are likely to build similar AI agents themselves in the future, this experience helping them make more thoughtful decisions about what kind of experiences the robot they construct should offer to the users (Eicher et al., 2018).

**Approach to this investigation**

This research is based on a literature review around several topics including the definition and current scope of VTAs, the emerging market space and potential opportunities of virtual tutor-related products, application and research of VTAs in an intelligent classroom, exploration of the future direction of VTAs, and policy and ethical challenges. It relies on multiple types of sources including academic sources, press releases, and policy reports. Press releases covered funding rounds for emerging technology companies in the broader education technology and virtual technology ecosystem. Multilateral organizations provided insightful coverage of the latest
developments in the education sector through annual reports. Given the evolving policy landscape surrounding education and tutoring in China, the literature review also scrutinized policy announcements from the Ministry of Education of the People’s Republic of China.

The initial research was complemented by thirteen interviews in total with fourteen different interviewees. All interviews with teachers were conducted for roughly an hour over video conferencing in Mandarin Chinese. The transcripts were then translated into English for the purpose of sharing research with the group and conducting analysis. To summarize the insights from these interviews, we categorized them according to a few topics: a description of the status quo of after school tutoring, the challenges faced by teachers, pain points where artificial intelligence can potentially help alleviate teacher’s responsibilities, and the interviewees’ concerns over the use of virtual teaching. The rest of the interviews were semi-structured and lasted between 30 and 60 minutes, during which we took notes. The insights from these interviews were summarized into the over-arching categories/themes in the following section, incorporating the findings from the teacher interviews. We also systematically asked our interviewees to share two or three key references on the topics we discussed together in order to refine our initial literature review. We interviewed nine academic practitioners and industry experts to get a better understanding of where the tech currently is, what is concretely being developed at this stage, and even the next steps and potential new features that future applications of virtual teaching assistants are likely to have. Five elementary and middle school teachers were interviewed in an effort to validate the potential for VTAs to help them, understand their pain points -especially for after school tutoring - and gather ethical, policy-related, and pedagogical objections or apprehensions concerning VTAs.

**Building and deploying VTAs**

*Positioning VTAs: Users and methods*

The greatest potential for VTAs lies in the menial tasks that take up teachers’ time yet could be automated by AI, but VTAs are not a viable alternative to the complexity, understanding, and empathy of human teachers who perform the monumental task of shaping the lives of young students on a daily basis. Given these limitations, we consider which actors should use VTAs, when they should be deployed during a school year, and which subjects these are best suited for.

There are strong incentives for individual students to seek out help from a VTA, given they can help increase their scores by 5-10 points on the gaokao, making them jump up a tier in higher education. Schools are in turn rewarded for improved performance by students on the gaokao (Pan et al., 2016). The incentives therefore align best for selling to schools and designing the VTA to include features relevant to both teachers and students, such as automated grading and personalized reports on student exam performance over time. According to our teacher interviewees, additional homework help, question-and-answer on the lessons given during the day, and self-study are tasks that students generally perform on their own, not with groups of other students. Products that assist students with self-study and adaptive learning have become very popular in recent years, especially up until DRP (Zhan, 2019). Furthermore, VTAs are not effective at assisting creative or collaborative tasks, so they are more well-suited to the tasks listed above. They are particularly useful in the context of review for major exams such as midterm and final tests for three reasons. In preparation for large exams, students can look back on historical reports of their performance in different quizzes and tests over the course of the semester to hone in on weak knowledge points. In this context, VTAs also help teachers grade the large amounts of exercises that students are using for review. Lastly, whilst VTAs are not a perfect solution for
online education, their personalized nature, ability to engage students, and integrated use in the curriculum make them appealing as an aid to online education.

Through engagement on the online platform, students can improve information retrieval through spaced practice (Johnson, 2018). Quizzing applications such as Kahoot and Quizlet have made this type of learning possible and effective. Interviewees suggested that material should be scaffolded in such a way that learners are given extra time to go through the content at their own pace. Since the teachers do not know what is going on in the student’s brain, it is important to ensure that online students have plenty of content and time to learn and assimilate. Further, it is not enough to simply grade students and give them their score: students should be taught how to do their own peer assessment.

At the same time, not all subjects are equally well-suited for VTAs; the language arts require more open-ended thinking (Flaherty, 2020), while the sciences generally require more closed-ended questions with discrete curriculum units. Humanities subjects struggle with online teaching and learning at scale, because it is difficult to replicate in-class small group discussions online. Automated grading is better suited to students in STEM courses, especially with recent breakthroughs in NLP such as GPT-3 that have enabled improved accuracy when grading open-ended questions. Erickson et al. (2020) assessed the accuracy of automated grading in open-ended questions within mathematics and found that all models showed at least 37% improvement over the baseline of pure chance. However, grading rubrics are still scripted, and artificial intelligence-driven programs are not necessarily able to catch the nuances that are desired as of now. English language learning is a good candidate for assistance from VTAs, because most of the natural language processing advances were first made in English, so there are relatively high amounts of standardized test questions in that language. For oral learning, an AI program can also judge pronunciation of English words and the intonation of sentences, such as the speech recognition software iFLYTEK.

**Pegadogical benefits of VTAs and product features**

For teachers, grading is the most challenging and effort-intensive part of the work, some spending up to 4 hours every day on correcting homework. At the same time, they experience pressure from the preparation of materials and curriculum for after school classes. This is even more problematic in schools where students of many levels are packed into a single class, because teachers struggle to assign homework appropriate for students’ relative levels while also teaching a standardized curriculum.

**Automated grading** The grading capability of VTAs for multiple choice questions is already mature, and the current progress in machine learning is directed towards natural language processing for grading open-ended questions, including short answers and fill-in-the-blank. In the context of coding, an ideal VTA solution should guide the student to the correct answer, by telling them which areas of the code are correct or not, rather than simply revealing the whole answer. At present, AI is not necessarily able to catch inefficiencies in the code, or attribute partial credit where students deserve it.

**Grading analysis** Beyond grading, the customized grading reports generated by VTAs can serve both teacher and student needs by providing the learner with an individualized “performance profile”. One area in NLP that bears great potential is calculating the distance of a student’s answer from a standardized set of answers created by the teacher, taking into account semantic accuracy, logical reasoning, and stylistic excellence. The VTA can take students’ answers and compare them against this baseline. In their study, Ye and Manoharan (2019) found that since all students’ work can be embedded into a continuous vector space, calculating the Euclidean distance between them is trivial. For example, a student’s essay can be used to
generate multiple embeddings – in theory, a VTA could learn multiple different models for embedding an essay’s semantics, ranging from style to logical reasoning. By synthesizing insights from students’ entire learning history, these reports can help them have a bird eye’s view of the issues associated with their performance over the course of a long semester. At the same time, teachers can better understand which students need help, and with which specific topics. This is particularly valuable given that those who fall below the average level of proficiency may feel overwhelmed by homework, too intimidated to participate in class discussions, and develop a sense of inadequacy over their abilities, whilst those significantly above the class average may be bored by lessons, disinterested in class discussions, and generally develop a sense of apathy toward their education.

**Customized curriculum** Customizing the curriculum and homework to all students, especially when there is a large level of variation, is a potential use case for after school VTAs. Whilst also seeing how the class is doing at an aggregate level, teachers can decide which lessons to prioritize, opening up new possibilities for lesson plans by avoiding to spend redundant time reinforcing a topic when it has already been taught and learned. Waikar (2020) describes a Stanford team building an AI-based platform to assist students with the process of navigating open-ended tasks such as coding. They proposed a “Super Teaching Assistant” that provided customized feedback by generating automated reports highlighting the challenges students faced on both a personal and class-wide basis. This applies particularly well in classes where curriculum units are discretely defined, enabling the VTAs to create personalized grading reports for students based on their performance in each unit. Singla et al. (2021) surveyed the existing literature on reinforcement learning (RL) for education, a branch of machine learning which seeks to optimize decisions made in sequential environments where partial information is revealed at each step and the situation evolves at each step, perhaps within a limited set of partially observed states. They found that RL is well-suited for the problem of curriculum planning and student improvement, since it generally proceeds in discrete steps over the course of a school year. This is illustrated in the example below for a maths class (Figure 1).

The customized score card could be used not only for exam results or semester-long assessments, but weekly practices or exercises. Teachers could have an input on how often the score cards are generated, depending on the frequency at which students are given exercises to practice.

![Customized report card for a student's performance on different topics for the math course of fall 2025.](image)

**Figure 1.** Customized report card for a student’s performance on different topics for the math course of fall 2025.
on their weaknesses, and the speed at which they progress. Customized lessons then allow teachers to teach students of multiple levels simultaneously by allowing the VTA to design lessons with questions of varying difficulty.

**Implementation**

In this section, we detail some of our findings surrounding the implementation of VTAs. In particular, this entails aspects of user interface, hosting VTAs on hardware, how to involve relevant stakeholders, the system design principles and human element involved.

User interface and user experience (UI/UX) is critical to adoption by students, because they must feel comfortable and encouraged without trading off the quality and rigor of the educational content. To this end, “intelligent interaction” can be used to inject human emotions in the process of interaction between robots and humans. Sentiment analysis and facial recognition are techniques that can help the VTA grasp both the pedagogical and psychological needs of students. In 2018, the TAL Education Group released the Magic Mirror, a brilliant emotion analysis product driven by technologies such as image recognition, speech recognition, and gesture recognition (Zhan, 2019). This product can help virtual teaching assistants understand students’ emotions and learning interests in real-time. Furthermore, research shows the benefits of the “persona effect”, a well-studied phenomenon in education (Lester et al., 1997), whereby an animated tutor in a dynamic educational situation can have beneficial effects on a student’s feelings towards their learning. This study suggests that a VTA may benefit from a moderate amount of anthropomorphization, whilst ultimately ensuring a balance of engagement and distraction.

When it comes to hardware, around 74% of children under 18 own their smartphones and regularly surf the internet on their devices (Wakefield, 2021). However, The Ministry of Education ruled in 2021 that children in China are not permitted to use their mobile phone at school without written consent from their parents (Wakefield, 2021), and individual schools also have their own guidelines. Instead, schools could purchase a number of laptops or tablets in bulk, and form a separate “computer laboratory” where students can reserve a slot after school or an entire class can reserve an hour for use. Mahalli et al. (2019) found that in a “station learning” system, students can rotate onto the laptops for part of the class period and spend the rest of the period collaborating with each other or working with the human teacher. The diversity of learning strategies that station rotation enables has been found to deepen student learning, improve interactivity, boost curriculum flexibility, and augment personalized learning. At the same time, this means that devices cannot be taken home for students to use in the remote learning scenario, but it remains the most effective model for hardware ownership given the current regulations in place and overall cost-effectiveness.

Administrators and technology developers may not have the know-how or experience to predict every need and use-case of VTAs, so it is critical to involve teachers, parents, and students in the implementation phase. Teachers have different needs based on the topic instructed and teaching experience. They should be allowed to raise opinions about desired features, offer suggested adjustments to existing technology, and hypothesize about the optimal rollout of the new program. Given the switch from private tutors to after-school tutoring in public schools, the doubts of parents regarding technology are understandably higher than usual. To assuage their concerns and gain their support, parents should have access to the content created by the VTAs, such as the personalized report cards and customized lessons plans. Furthermore, Wang and Xing (2018) argue that teens whose parents were more involved in their technology usage developed better digital etiquette, privacy habits, and attitudes toward social media. Finally, since students lie at the core of the VTA effort, their wellbeing should be closely monitored throughout the full deployment process. This can be achieved through informal conversations, formal surveys, and
special hotlines in cases of great distress or discomfort, and later, ensuring constant feedback loops even after implementation.

**Critical perspectives on VTAs**

Challenges fall into three major categories: technical, pedagogical, and ethical. Technical problems are associated with the implementation of VTAs in classrooms and the technological barriers associated with deploying AI in a dynamic setting like education. Pedagogically speaking, VTAs must integrate effectively with teachers and boost student learning without distracting them from the core curriculum. Lastly, ethical challenges involve issues with privacy and the emotional impact of VTAs.

**Technical challenges**

A core challenge facing developers of VTAs is the acquisition of sufficient, high-quality data. The large language models mentioned earlier in the report depend on an enormous amount of input-output pairs, with the input being the essay or short answer, and the output being the grade or teacher’s comments.

Since artificial intelligence relies on historical data to train its algorithms, the questions generated and recommended to students may not correspond to the topics and skills they need to succeed in the dynamic educational environment. Teaching also involves seeing the thought process of students, not just their final answer. If a VTA could help teachers grade homework, it is crucial for the VTA to help track those thought processes of students for teachers. This is particularly true for teachers for whom correcting homework manually is a cornerstone of their approach to personalizing their teaching style. Furthermore, the gaokao exam was reformed to increase the amount of reading comprehension and writing analysis questions, which are challenging for AI to grade accurately without human assistance. However, AI interpretability in education is a burgeoning field (Belinkov et al., 2020). The rise of attention in models like the Transformer and BERT allow models to pinpoint parts of sentences that are the most important for a prediction, which bear great potential for automated grading. In the meantime, if an AI product incorrectly grades a question and then requires human intervention, the time wasted negates the benefit of using a virtual teaching assistant at all.

Finally, VTAs may run into resource limitations given that most schools in China do not allow cellphones in classrooms.

**Pedagogical challenges**

Since every class session has new content every day, new topics, new homework, and extensions on existing topics, a VTA may have trouble achieving the granularity needed for daily tutoring—this suggests that VTAs may play a better role as occasional assistants rather than daily technology. If students receive additional homework for topics they do not understand, it will increase the burden and actually deviate from the original intention of DRP. Furthermore, automated question generation may not greatly ease the teaching burden, because teachers fear the questions might not be substantial or well written enough. Targeted homework guidance by VTAs may thus only be suitable for certain stages of the learning process, such as the review sessions near the end of a semester rather than at the outset of the semester.

From a teaching perspective, virtual teaching assistants might stifle their creativity by hindering their initiative, because a VTA restricted to a certain question bank and a style of teaching by computer programs may not complement a teacher who wants to use creative ways of teaching,
such as student skits, artistic assignments, or peer interaction. Since artificial intelligence can only grade close-ended multiple choice questions and fill-in-the-blank questions, it struggles to communicate lessons involving proper relationships between people, referred to as “moral education”. This is particularly critical for students in hollowed out villages, where parents may live only part time at home or, in the worst case, not at all (Liu et al., 2010), in which case educators play a key role.

Finally, some parents fear that if VTAs are taking the place of human teachers, it is because schools are attempting cost-cutting measures and therefore not offering optimal quality education.

**Ethical challenges**

It is challenging to classify and assess how a VTA performs on the more intrinsic, relationship-focused aspects of a teacher-student relationship. This relationship possesses a friendship dimension, both instrumentally and intrinsically valuable, which is based on variables such as authenticity and non-deception. Some researchers have argued though that deception could be beneficial if its purpose is well defined, such as modeling certain growth behaviors, for example by making a robot intentionally make a mistake, and make it think to itself that it should improve next time.

At present, there are fears that we live in an outcome-driven society which will define an edtech tutoring application as being successful if it helps a child acquire a set of skills, which fails to fulfill the needed emotional connection for students. If a benefit of these robots is cost-effectiveness, and assuming that we want to improve learning and ameliorate our education systems, some researchers suggest it would be more sensible to pour more money into the field, not less.

**Three key suggestions for VTA technology and adoption**

Drawing from our prior research and the analysis of interviews, we have devised a series of recommendations for stakeholders to consider as they develop and distribute VTA technology.

**Applications**

VTAs should be implemented for tasks that are more automatable, such as grading, starting with binary question types such as true/false and multiple choice, then later expanding to fill-in-the-blank questions. As for subjects of focus, maths and science, as well as English language, are the best candidates, and we recommend that the VTA begins its operations on subjects such as math and science.

One of the most exciting parts of a VTA is its ability to collect a volume of data about both the class as a whole and individual students. To this end, VTA technology is designed to give teachers a dashboard view of the performance of their class so that they can tailor their future teachings accordingly. The VTA should also provide feedback on an individual basis so that the teacher can provide personalized help to students. Further, it should be able to automatically assist them through providing real-time feedback and detailed explanations for students based on their responses to questions. For example Aidam, Xuebajun’s education robot, can roughly imitate the perception, memory, cognition and analysis capabilities of the human brain. By establishing a knowledge base, the whole process of association, assessment and decision-making can address all questions typically asked in middle school, so that the machine can grade papers, personalize guidance and homework, and improve course preparation (Lv, 2017).
Technology development and rollout

With the goals of equity and access, a technology device rental program that allows schools, regardless of financial status, to be able to access the devices on which VTA technology runs, appears to be the best option. As with all technology, it is not expected that the release of these applications in schools will be perfectly smooth. However, student learning and well-being, as well as the well-being of teachers, is at stake with the launch of VTA technology. To this end, teachers must be involved throughout the development and testing process of its technology. The first year of launching a VTA technology is a trial run where they are continually seeking and receiving input from both students and teachers about the effectiveness of its technology. Importantly, the evaluation data on the success of the technology should be focused not only on how it helps students academic performance, but also how it impacts student mental and emotional well-being. Relatedly, teachers should continually be a part of the development process of VTA technology, such as allowing them to share their VTA-reliant apps with other teachers and also providing input to the question bank-related elements of the VTA.

Education around VTAs

While VTAs hold many exciting promises in terms of increasing the personalization of the student learning experience, they also can inspire new anxieties in students and their families. Relatedly, developers should devise materials which explains to students and families how the VTA is being used and how it can benefit their education. Importantly, it must be emphasized to teachers that they are not being replaced. Rather, it should be explained that VTA technology is making a teacher more efficient, helping them better understand their class on both a macro and individual student basis, and subsequently assists them in providing detailed feedback and personalized lesson plans to students of different abilities.

Conclusion

VTAs will likely be a significant part of the future teaching and classroom experience for students across a variety of grade levels. Compared to other existing forms of virtual e-learning such as Learning Management Systems, VTAs provide a deeper level of interaction with students. They play a key role in student engagement beyond merely tracking their progress on defined metrics. However, this study reveals that it is vital that VTAs are not viewed as an end-to-end solution to the burdens placed on teachers. There are applications such as grading, especially that with binary answers, where VTAs can help, but they cannot fully replace the role of a teacher and still ensure student well-being. With VTAs, teachers can also understand their classroom performance, allowing them to customize lesson plans and assignments for their students. To ensure VTAs are deployed in a way that addresses stakeholder needs, a trial-and error approach, supplemented by thorough education and complementary resources around the use of this technology, should be adopted.

While considerable research has been done around what may work best as it relates to a VTA application, there are no guarantees that it will work in practice. It is important that teachers are prepared to abandon the use of parts of the VTA, or perhaps even the VTA in full, if it is deemed that some parts are not working well. Technology development can sometimes wait for iterations, but student well-being cannot.

Like many applications of technology in education, the solution is hybrid, and adapting to educational contexts and cultures is key. Given the maturity of English automated grading artificial intelligence compared to other languages, VTAs may not integrate as smoothly in all
environments. Furthermore, countries that have already heavily incorporated technology into education curricula might be better prepared to integrate AI-driven tool like VTAs, highlighting the need to consider the local context before deploying such products.

Acknowledgements

We would like to thank Dr. David Pan, our advisor and mentor at Schwarzman College, as well as the fourteen interviewees. Additionally, we would like to thank Sam Huang and Yuan Xin at Tencent for providing guidance, resources, and the motivation for our project throughout the year.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Note

1. We reached out to these specific individuals because we came across their names in our literature review, they were recommended to us, or because we thought they would provide valuable information based on our preliminary readings. Teachers are in Beijing and Zhejiang, and have been contacted through previous connections with the schools they work at. The majority of the academic professionals and industry experts work in the United States (Massachusetts Institute of Technology, Athabasca University) and United Kingdom (University of Oxford), others in China (Tsinghua University) and Europe (Technische Universität Berlin). They have been contacted through connections at the institutions they work at, or cold e-mailed.

References

Aiken, R. M., & Epstein, R. G. (2000). Ethical guidelines for AI in education: Starting a conversation. International Journal of Artificial Intelligence in Education, 11, 163–176. https://doi.org/10.1126/scirobotics.aat5954

Anderson, J. (2017). Asians spend seven times as much as Americans on tutoring to give their kids an edge. Quartz. https://qz.com/970130/asians-spend-15-of-their-family-income-on-extra-education-and-tutoring-for-kids-americans-spend-it-on-cars-and-gas/

Belinkov, Y., Gehrmann, S., & Pavlick, E. (2020). Interpretability and analysis in neural nlp. In Proceedings of the 58th annual meeting of the association for computational linguistics: Tutorial abstracts (pp. 1–5). Online, 2020, July 5. https://doi.org/10.18653/v1/2020.acl-tutorials.1

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. Science Robotics, 3(21), Article eaat5954. https://doi.org/10.1126/scirobotics.aat5954

Carlana, M., & La Ferrara, E. (2021). Apart but connected: Online tutoring and student outcomes during the covid-19 pandemic. Edworkingpaper no. 21-350. Annenberg Institute for School Reform at Brown University. https://www.edworkingpapers.com/ai21-350

Dandan, N. (2021). A is for anxiety: Tutoring clampdown tests China’s parents. SixthTone. https://www.sixthtone.com/news/1008347/a-is-for-anxiety-tutoring-clampdown-tests-chinas-parents

Eicher, B., Polepeddi, L., & Goel, A. (2018). Jill watson doesn’t care if you’re pregnant: Grounding ai ethics in empirical studies. In Proceedings of the 2018 AAAI/ACM conference on AI, ethics, and society (pp. 88–94). New Orleans, LA, USA. 2018, February 2 - 3. https://doi.org/10.1145/3278721.3278760
Erickson, J. A., Botelho, A. F., McAteer, S., Varatharaj, A., & Heffernan, N. T. (2020). The automated grading of student open responses in mathematics. In Proceedings of the tenth international conference on learning analytics & knowledge (pp. 615–624). Frankfurt, Germany. 2020, March 23 - 27. https://doi.org/10.1145/3375462.3375523

Flaherty, C. (2020). ‘As human as possible’. Inside Higher Ed. https://www.insidehighered.com/news/2020/03/16/suddenly-trying-teach-humanities-courses-online

Furness, D. (2020). How AI will completely transform education. Digital Trends. https://www.digitaltrends.com/computing/how-ai-is-changing-education/

Johnson, R. (2018). Supporting retrieval practice with quizzing technology. https://techandcurriculum.pressbooks.com/chapter/supporting-retrieval-practice/

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., & Bhogal, R. S. (1997). The persona effect: Affective impact of animated pedagogical agents. In Proceedings of the ACM SIGCHI conference on human factors in computing systems (pp. 359–366). Atlanta, Georgia, USA. 1997, March 22 - 27. https://doi.org/10.1145/258549.258797

Liu, Y., Liu, Y., Chen, Y., & Long, H. (2010). The process and driving forces of rural hollowing in China under rapid urbanization. Journal of Geographical Sciences, 20(6), 876–888. https://doi.org/10.1007/s11442-010-0817-2

Lv, M. (2017). Jiqiren Ye Gaokao? Xuebajun Jiaoyu Jiqiren Aidam 134 Fen Naxia Gaokao Shuxueti. Lie Yun Wang. https://www.lieyunwang.com/archives/320941

MahalliNurkamto, J., Mujiyanto, J., & Yuliasri, I. (2019). The implementation of station rotation and flipped classroom models of blended learning in efl learning. English Language Teaching, 12(12), 23–29. https://doi.org/10.5539/elt.v12n12p23

Nickow, A., Oreopoulos, P., & Quan, V. (2020). The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3644077

Patel, K., Mehta, D., Mistry, C., Gupta, R., Tanwar, S., Kumar, N., & Alazab, M. (2020). Facial sentiment analysis using AI techniques: State-of-the-art, taxonomies, and challenges. IEEE Access, 8, 90495–90519. https://doi.org/10.1109/ACCESS.2020.2993803

Singla, A., Rafferty, A. N., Radanovic, G., & Heffernan, N. T. (2021). Reinforcement learning for education: Opportunities and challenges. (arXiv:2107.08828). arXiv. http://arxiv.org/abs/2107.08828

Song, M. (2022). Under the implementation of double reduction policy: Relationships between teachers and parents. In 2021 International conference on education, language and art (ICELA 2021), Sanya, China, 26 - 28 November 2021. https://doi.org/10.2991/assehr.k.220131.146

Waikar, S. (2020). How an AI-based “super teaching assistant” could revolutionize learning. Stanford University School of Engineering. https://engineering.stanford.edu/magazine/article/how-ai-based-super-teaching-assistant-could-revolutionize-learning
Wakefield, J. (2021). *China bans children from using mobile phones at school*. BBC News. https://www.bbc.com/news/technology-55902778

Wang, Q., Saha, K., Gregori, E., Joyner, D., & Goel, A. (2021). Towards mutual theory of mind in human-ai interaction: How language reflects what students perceive about a virtual teaching assistant. In Proceedings of the 2021 CHI conference on human factors in computing systems (pp. 1–14). Yokohama, Japan. 2021, May 8 - 13. https://doi.org/10.1145/3411764.3445645

Wang, X., & Xing, W. (2018). Exploring the influence of parental involvement and socioeconomic status on teen digital citizenship: A path modeling approach. *Journal of Educational Technology & Society*, 21(1), 186–199. http://www.jstor.org/stable/26273879

Yang, Y., Luo, W., & Tong, H. (2019). Jiuy “renxing jiegou” lilun de AI zhujiao xitong moxing yanjiu - hangye zhishi fuwu pingtai [study on the model of AI teaching assistant system based on theory of “human nature structure”]. *Dianhua Jiaoyu Yanjiu*, 40, 12–20. http://cglhub.com/auto/db/detail.aspx?db=190716&rid=75527219&agfi=0&cls=0&uni=True&cid=0&showgp=False&prec=False&md=54&pd=210&msd=54&psd=210&mdd=54&pdd=210&count=10&reds=%E6%95%99%E8%82%B2%3B%E5%88%9B%E5%AE%A2

Ye, X., & Manoharan, S. (2019). Providing automated grading and personalized feedback. In Proceedings of the international conference on artificial intelligence, information processing and cloud computing - AIIPCC ’19 (pp. 1–5). Sanya, China. 2019, December 19 - 21. https://doi.org/10.1145/3371425.3371453

Zhan, X. Q. (2019). The market research of AI online education industry of China in 2019 (Leadleo Industry Report 19RI0796). *Wind Database*. 