

ITS, The End of the World as We Know It: Transitioning AIED into a Service-Oriented Ecosystem

Benjamin D. Nye¹

Published online: 2 February 2016

© International Artificial Intelligence in Education Society 2016

Abstract Advanced learning technologies are reaching a new phase of their evolution where they are finally entering mainstream educational contexts, with persistent user bases. However, as AIED scales, it will need to follow recent trends in service-oriented and ubiquitous computing: breaking AIED platforms into distinct services that can be composed for different platforms (web, mobile, etc.) and distributed across multiple systems. This will represent a move from learning platforms to an ecosystem of interacting learning tools. Such tools will enable new opportunities for both user-adaptation and experimentation. Traditional macro-adaptation (problem selection) and step-based adaptation (hints and feedback) will be extended by meta-adaptation (adaptive system selection) and micro-adaptation (event-level optimization). The existence of persistent and widely-used systems will also support new paradigms for experimentation in education, allowing researchers to understand interactions and boundary conditions for learning principles. New central research questions for the field will also need to be answered due to these changes in the AIED landscape.

Keywords Artificial intelligence · Education · Meta-adaptation · Micro-adaptation · Distributed systems · Intelligent tutoring systems · Social computing · Adaptive sampling · Reinforcement learning · Transfer learning · Ontologies

Introduction

Initial efforts to bring learning technology into schools faced hardware hurdles, such as insufficient computing resources. Later efforts encountered serious barriers related to matching technology to teachers' beliefs, pedagogy, and resource constraints. While all of these barriers are still relevant, learning technology is endemic in higher education

✉ Benjamin D. Nye
benjamin.nye@gmail.com

¹ Institute for Creative Technologies, University of Southern California, 12015 Waterfront Dr., Playa Vista, CA, USA

and has made significant footholds in K-12 schools, with estimates of 25–30 % of science classes using technology as early as 2012 (Banilower et al., 2013). Correspondingly, an influx of investment into educational technology has occurred, with online learning doubling from a \$50b industry to a \$107b industry in only three years (Monsalve, 2014).

Future barriers will not be about getting learning technology into schools: they will be about competing, integrating, and collaborating with technologies already in schools. This is not an idle speculation, as it is already occurring. In a recent multi-year efficacy study to evaluate a major adaptive learning system, some teachers started using grant-purchased computers to use other math software as well (Craig et al., 2013). After working for many years to get teachers to use technology, the point may come where they are using so many technologies that it is difficult to evaluate an intervention in isolation.

Some research-based artificial intelligence in education (AIED) technologies have already grown significant user bases, with notable examples that include the Cognitive Tutor (Ritter et al., 2007), ALEKS (Falmagne et al., 2013), and ASSISTments (Heffernan et al., 2006). Traditionally non-adaptive systems with large user bases, such as Khan Academy and EdX, have also started to add basic adaptive learning and other intelligent features (Khan Academy, 2015; Siemens, 2013).

Large-scale online platforms are not just the future of learning, but they are also the future of research. Traditional AIED studies have been limited to dozens to hundreds of participants, sometimes just for a single session. While such studies will remain important for isolating new learning principles and collecting rich subject data (e.g., biometrics), large-scale platforms could be used to run continuously-randomized trials across thousands of participants that vary dozens or even hundreds of parameters (Mostow and Beck 2006; Liu et al., 2014). Even for AIED work not based on such platforms, it is increasingly feasible to “plug in” to another system, with certain systems serving as active testbeds for 3rd-party experiments (e.g., ASSISTments and EdX).

The difference is qualitative: rather than being limited to exploring a handful of factors independently, it will be possible to explore the relative importance of different learning principles in different contexts and combinations. In many respects, this means not just a change to the systems, but to the kinds of scientific questions that can and will be studied. These opportunities raise new research problems for the field of AIED. A few areas related areas will reshape educational research: Distributed and Ubiquitous Intelligent Tutoring Systems (ITS), Four-Loop User Adaptation, AI-Controlled Experimental Sampling, and Semantic Messaging. Some new frontiers in each of these areas will be discussed.

Distributed and Ubiquitous AIED

As implied by the title, AIED technologies are approaching a juncture where many systems will be splitting up into an ecosystem of reusable infrastructure and platforms. The next generation of services will be composed of these services, which may be hosted across many different servers or institutions. More specifically, we may be reaching the end of the traditional four-component ITS architecture with four modules: Domain, Pedagogy, Student, and Communication (Woolf, 2010). While the functions

of all these modules will still be necessary, there is no reason to think that any given ITS must *contain* all these components, in the sense of building them, controlling them, or owning them. The future for ITS may be to blow them up so that each piece can be used as a web-service for many different learning systems.

With respect to other online technologies, learning technology is already behind. On even a basic blog site, a user can often log in using one of five services (e.g., Google, Facebook), view adaptively-selected ads delivered by cloud-based web services that track users across multiple sites, embed media from anywhere on the internet, and meaningfully interact with the site on almost any device (mobiles, tablets, PC). In short, most web applications integrate and interact with many other web services, allowing them to be rapidly designed with robust functionality and data that no single application would be able to develop and maintain.

The underlying drive and principles for breaking AIED down into components and services is hardly new. Work on this area started over two decades ago (Brusilovsky, 1995; Futersack & Labat, 1992) and lack of progress was noted not long after (Roschelle & Kaput, 1996). The principles discussed at that time for integrating distributed systems still remain valid today (e.g., whiteboard architectures, agent-oriented designs). The failure to embrace those designs was likely due to a lack of absolute need: the use of educational technology was low, there were not many platforms to integrate, and architectures for distributed web-based services were still in their infancy. By comparison, in the modern day, a high school student could take a tutorial and set up an Amazon micro-service within a day.

From the standpoint of AIED now, moving toward distributed systems is an existential necessity. Without pooling capabilities or sharing components, serious academic research into educational technologies may be boxed out or surpassed by the capabilities of off-the-shelf systems (which may or may not report generalizable findings). Academic institutions and research-active commercial systems should be motivated to share and combine technologies to build more effective and widely-used learning technology. This model of collaborative component design stands alone in making platforms that co-exist with major commercial endeavors, such as web-browsers (FireFox), operating systems (Linux), and statistical packages (R; R Core Team, 2013). Moreover, service-oriented computing allows for a mixture of free research applications and commercial licensing of the same technologies.

The benefits of moving toward service-oriented AIED will be substantial. First, they should enable AIED research to deeply specialize, while remaining widely applicable due to the ability to plug in to other platforms with large and sustained user bases. In such an ecosystem, user adaptation will be free to expand beyond the canonical inner loop and outer loop model (VanLehn, 2006). Composing and coordinating specialized AIED services will also demand greater standardization and focus on data sharing between systems. Approaches for generating ITS tasks based on knowledge bases should also be complementary (e.g., El-Sheikh & Sticklen, 2002), since these could potentially plug in to existing services for advanced functionality. While this process may be painful initially, standards for integrating data across multiple systems would enable the development of powerful adaptation, analytics, and reporting functionality that would greatly reduce barriers for developing AIED technology and studying its effects on learners.

Four-Loops: Above Outer Loops and Under Inner Loops

One implication of scaling up AIED and moving beyond the standard four-component ITS model is that adaptation to users may become prevalent at grain sizes larger and smaller than traditional ITS. VanLehn (2006) framed the adaption from tutoring systems as consisting of an outer loop (selecting problems) and an inner loop (providing help and feedback on specific problem steps). These are often referred to as “macro-adaptivity” and “step-based adaptivity.” However, recent developments have shown the first steps toward “meta-adaptivity,” where the system adapts to the user by shifting the learner to an entirely different ITS system (which may then adapt to the user differently). Likewise, research on “micro-adaptivity” has looked at the benefits for using data to fine-tune interactions below the problem step level (e.g., keystroke-level inputs, emotion detection, presentation modes or timing of feedback; Graesser, 2011). This implies a four-loop model for user adaptation, as shown in Fig. 1. To note, there may be other loops of interest and VanLehn (2015) has likewise more recently framed ITS behavior more generally as regulatory loops. However, from a design standpoint, the boundaries between ITS systems, learning activities, and steps of a learning activity are likely to remain key decision points for adaptive systems (i.e., four natural loops).

Meta-Adaptation: Handoffs Between Systems

Meta-adaptation has only become possible recently, due to increasing use and maturity of AIED technology. In the past, learning technologies such as ITS were trapped in sandboxes with no interaction. Due to service-oriented approaches, systems have taken the first steps toward real-time handoffs of users between systems. For example, in the

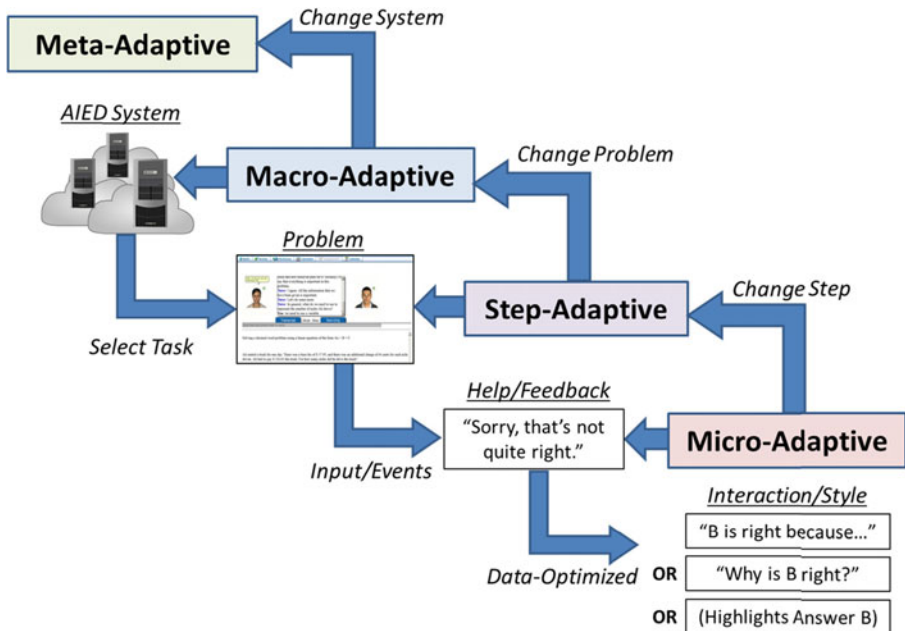


Fig. 1 Four-Loop User Adaptation

recent Office of Naval Research STEM Grand Challenge, two out of four teams integrated multiple established adaptive learning systems: Wayang Outpost with ASSISTments (Arroyo et al., 2006; Heffernan et al., 2006) and AutoTutor with ALEKS (Nye et al. 2015). Other integration efforts are also underway as part of the Army Research Lab (ARL) Generalized Intelligent Framework for Tutoring (GIFT) architecture, which is built to integrate external systems (Sottolare et al., 2012) and version of AutoTutor has also been integrated with GIFT.

These initial integrations represent the first steps toward meta-adaptation: transferring the learner between different systems based on their needs and performance. This type of adaptation would allow learners to benefit from the complementary strengths of multiple systems. For example, learners that benefit most from animated agents might be sent to systems such agents (i.e., trait-based adaptation). Alternatively, different types of learning impasses or knowledge deficiencies may respond best to learning activities in different systems (i.e., state-based adaptation). One problem that this approach might mitigate is the issue of wheel spinning, where an adaptive system detects that it cannot serve the learner's current needs (Beck & Gong, 2013). As such, one factor for switching between systems might be the estimated learning curves for different knowledge components, which could identify plateaus (e.g., wheel spinning) or comparative advantages (e.g., systems that specialize in certain topics or stages of learning).

Meta-adaptation can also mean referring the learner to a human instructor, tutor, or peer. This intersects with computer-supported collaborative learning (CSCL), which has a long history of trying to optimize how and when different educational stakeholders should interact (e.g., learners, teachers, parents; Kirschner & Erkens 2013; Stahl, 2010). CSCL research highlights the importance of groups of learners, both for motivation (e.g., leveraging friendships, authority roles, and other social ties) and as learning tasks (e.g., guided discussions). An obvious implication to this is that learners might not switch between systems in isolation, but might instead might transition in cohorts, switch systems to follow friends, or directed to systems daily by an instructor (e.g., web-based homework). While a review of different CSCL mechanisms would be broader than the scope of this paper, research topics such as when to give guidance (e.g., alert a tutor or instructor) and how to build and maintain effective learning teams will be highly relevant. These mechanisms would also be particularly relevant for any possible meta-adaptation for self-directed learning, where teams of learners might need to be emergent rather than prescribed by a school.

Some intuitions from CSCL are also relevant to the process of meta-adaptation, in that an effective combination of learning technologies must work similarly to a human learning team as described by Kirschner & Erkens (2013, pp. 4): “mutual performance monitoring (keeping track of each other's task activities), back-up behavior (backing up other member's failures) and adaptability (ability to adjust task or team strategies).” In intelligent systems, these three measures are might framed as sharing data, averting wheel spinning (i.e., the failure of a system to help a learner), and adaptively coordinating multiple systems. Moreover, this also implies that, much as introducing a new learner to a team involves these costs, introducing learners to a new system, activity, or interface has similar costs. This means that the gains from system transitions need to exceed these overhead costs. As such, conceptualizing an ITS as a new “member” of a CSCL team of learners can uncover potential pitfalls (e.g., when learning an interface

exceeds the benefits for faster learning in a new system) and may be one useful way to frame meta-adaptivity until more specialized theories emerge.

In general, meta-adaptation involves passing students and knowledge between different adaptive learning contexts (both AI-based and human). Based on that view, it is important to consider how meta-adaptation currently occurs. For example, in traditional classes, teachers are the meta-adaptive controller: they choose between systems and student interactions, apply them during the curriculum, and try to learn about how and when they help students. Computer-based meta-adaptation (e.g., data trends, recommendations) for teacher-support might be a natural advance, and would also allow modeling how teachers make their decisions for choosing systems (e.g., a living “What Works Clearinghouse” for interleaving multiple systems in a course).

Meta-adaptation is the maximum possible grain size, which makes it somewhat different from standard adaptation because users are transferred to an entirely different system. This type of adaptation likely requires either distributed adaptation or brokered adaptation. Distributed adaptation would involve individual systems deciding when to refer a learner to a different system and possibly trusting the other system to transfer the student back when appropriate. This would be analogous to doctors in a hospital, who rely on networks of specialists who share charts and know enough to make an appropriate referral, but may use their own judgment about when and how they make referrals. On the converse, brokered adaptation would require a new type of service whose purpose is to monitor student learning across all systems (i.e., a student model integrator) and make suggestions for appropriate handoffs. This service would be consulted by each participating AIED system, probably as part of their outer loop. In the long term, such a broker may be an important service, because it could help optimize handoffs and ensure that students are transferred appropriately. Such brokers might also play a role for learners to manage their data and privacy settings. Other models for coordinating handoffs might also emerge over time.

Micro-Adaptation: Data-Optimization and Event Streams

In addition to adaptivity at the largest grain size (selecting systems), research on the smallest grain sizes (micro-adaptation) is also an important future area. Micro-adaptation involves optimizing for and responding to the smallest level of interactions, even those that are not associated with a traditional user input on a problem step. For anything but simple experiments, this type of optimization and adaptation is too fine-grained and labor-intensive to perform by hand at scale, meaning that it will need to rely on data-driven optimizations such as reinforcement learning. Chi et al. (2014) used reinforcement learning to optimize dialog-based ITS interactions in the Cordillera system for Physics, which showed potential gains of up to 1σ over poorly-optimized dialog or no dialog. Dragon Box has taken a related approach by optimizing for low-level user interface and click-level data, by applying trace-based models to find efficient paths for learning behavior and associated system responses (Andersen et al., 2013).

These lines of research represent the tip of the iceberg for opportunities for micro-adaptation. A variety of low-level data streams have not yet been leveraged. Continuous sensor data, such as emotion sensors or speech input waveforms, may present rich opportunities for exploring fine-grained user-adaptation based on algorithmic

exploration of possible response patterns. Low-level user interface optimization may also help improve learning, such as human-computer interaction design or keystroke-level events or mouse-over actions (i.e., self-optimizing interfaces).

Both the strength and the drawback of micro-optimization is that it will tightly fit the specific user interface or content (even down to specific words in text descriptions). Optimizing for a particular presentation of a problem can lead to learning efficiency gains by emphasizing parts that are salient to learning from that specific case, while skipping or downplaying other features. However, micro-level optimization will likely suffer from versioning issues (e.g., changes to small problem elements potentially invalidating prior data and policies) and also transferability issues (e.g., an optimized case not transferring well from a desktop to a mobile context).

Solutions to this problem are non-trivial and will require serious advances for both machine learning and knowledge representation. Services for micro-adaptation can be thought of as a few types: ones that model relatively universal patterns (e.g., emotion detectors based on biometrics such as pupil dilation) and ones that model context-specific interactions that require burn-in data before they work effectively (e.g., most natural language classifiers). At present, many models for micro-optimization (e.g., reinforcement learning) have not transferred well to new tasks or contexts, even for the same domain (Chi et al., 2014). Research on how to weight the relevance of prior data will be required to address issues related to altered problems or new contexts (e.g., mobile devices, classroom vs. home, different cultural contexts). This contains three problems that are only partially-solved: how to estimate the scope over which model components can be generalized (e.g., multi-level cross-validation), how to integrate context-specific evidence into a generalized model, and how much to weigh evidence which differs along different dimensions of context (e.g., learner, system, task, problem, age of the evidence). With that said, even in the pessimistic cases where micro-level adaptation only works for a specific version of a specific problem, if the number of sessions needed to collect useful burn-in data is fairly small (e.g., less than one hundred), there could still be strong benefits for many systems.

AI-Controlled Experimental Sampling

Techniques for micro-adaptation may also reshape experimental methods. Artificial intelligence can play a major role in the experimental process itself, which is a type of efficient search problem. Educational data mining research has already started looking at dynamically assigning subjects to different learning conditions based on multi-armed bandit models (Liu et al., 2014). Multi-armed bandit models assume that each treatment condition is like a slot machine with different payout distributions (e.g., student learning gains). Likewise, cognitive experiments to test different models have used adaptive design optimization to search the space more efficiently (Cavagnaro et al., 2010). Adaptive sampling methods are common in medical research, where it is important to stop treatments that show harms or a consistent lack of benefit. For intelligent systems, they can be used to explore new strategies, while pruning ineffective ones.

The field is only taking its first baby steps for these types of experimental designs. Fundamental research is needed to frame and solve efficient-search problems present in AIED experiments. Based on varying different parameters and interactions in the

learning experience, learning environments can search for interpretable models that predict learning gains. In the long term, models for automated experimentation may even allow comparing the effectiveness of different services or content modules, by randomly selecting them from open repositories of content.

The most difficult aspect of this problem is likely to be the interpretability. While multi-arm bandit models can be calibrated to offer clear statistical significance levels between conditions, models that traverse the pedagogical strategy space are often too granular to allow for much generalization. For example, some popular models for large learning environment focus on efficient paths or traces of learning behavior and associated system responses (Andersen et al., 2013). Unfortunately, these models are often not easily generalizable: they may capture issues tied to the specific system or may tailor instruction to specific problems so tightly that it is difficult to infer theoretical implications (Chi et al., 2014). However, other approaches focusing on model comparison (e.g., Cavagnaro et al., 2010) avoid these pitfalls, but with the limitation that they require pre-existing hypothesized models to compare. Approaches that lie between these two extremes are likely to be most productive (i.e., automating the process of theoretically-guided exploration), but such methods are not yet mature.

New techniques are needed that can automatically explore the space of pedagogical designs, but that can also output interpretable statistics that are grounded in theories and concepts that can be compared across systems. This is a serious challenge that probably lacks a general algorithmic solution. Instead, such mappings will probably be determined by the constraints of learning and educational processes. A second major challenge is the issue of integrating expert knowledge with statistically-sampled information. Commonly, expert knowledge is used to initially design a system (e.g., human-defined knowledge prerequisites), which is later replaced by a statistically-inferred model after enough data is collected. However, in an ideal world, these types of heterogeneous data would be gracefully integrated (e.g., treating expert knowledge as Bayesian prior weights). Future research in AIED will need to identify where this sort of expert/statistical hybrid modeling is needed, and match these problems with techniques from fields of AI and data modeling that specialize in these issues.

Ultimately, a goal of this work should be to blur the lines between theory and practice by building systems that can both report and consume theoretically-relevant findings. It should be possible to build general service-oriented experimental frameworks that help determine a policy to determine the assignment of independent variables for an experimental participant, based on a space of possible experimental variables and data on results from those variables. By assigning theoretical constructs to the input and output variables for such an optimization, this approach would allow directly comparing theories and also for exploring interactions between theoretically-grounded constructs. The payoff for such research would be much richer theory: rather than merely causal connections, efficient exploration could be used to find relative strengths, function approximations, and other relationships.

Semantic Messaging: Sharing Components and Data

To share technology effectively, AIED must move toward open standards for sharing data both after-the-fact (i.e., repositories) and also in real-time (i.e., plug-in architectures). The

first steps in these directions have already been taken. Two notable data repository projects with strong AIED roots exist: the Pittsburg Science for Learning Center (PSLC) DataShop (Koedinger et al., 2010) and the Advanced Distributed Learning (ADL) xAPI standards for messaging and learning record stores (LRS; Murray & Silvers, 2013). The IMS Global Specifications are also a move in this direction (IMS Global, 2015). Notably, these standards focus primarily on what happened (e.g. event streams), rather describing how the learning task works (e.g., metadata classifications).

Due to solid protocols in messaging technologies, the technical process of exchanging data between systems at runtime is not onerous. The larger issue is for a receiving system to actually apply that data usefully (e.g., understand what it means). Hidden beneath this issue is a complex ontology alignment problem. In short, each learning technology frames its experiences differently. When these experiences and events are sent off to some other system, the designers of each system need to agree about what different semantics mean. For example, one system may say a student has “Completed” an exercise if they viewed it. Another might only mark it as “Completed” if the learner achieved a passing grade on it. These have very different practical implications. Likewise, the subparts of a complex activity may be segmented differently (e.g., different theories about the number of academically-relevant emotions). While efforts have been made to work toward standards, this seldom solves the problem: the issue with standards is that there tends to be so many of them.

So then, ontology development must play a key role for the future of ITS interoperability. These standards must first cover learner data (e.g., events, assessments), but in order to leverage theoretical principles, should eventually cover different types of learning tasks, pedagogical theories, and subject matter domains (Mizoguchi & Bourdeau, 2000). There are multiple ways that this might occur. Assuming the number of standards is countable, it would be sufficient to have an occasional up-front investment to develop and update explicit mappings between ontologies by hand. While this is low-tech, it works when the number of terms is fairly small. For larger ontologies of AIED behavior and events, it may be possible to align ontologies by applying both coding systems to a shared task (e.g., build benchmark tasks that are then marked up with messages derived from each ontology).

By collecting data on messages from benchmark tasks, it may be possible to automate much of the alignment between ontologies, particularly for key aspects such as assessment. Research on Semantic Web technologies is also very active, and may offer other effective solutions to issues of ontology matching and alignment (Shvaiko & Euzenat, 2013). The final approach is to simply live without standards and allow the growth of a folksonomy: common terms that are frequently used. These terms can then become suggested labels, with tools that make their use more convenient and prevalent. The one approach that should *not* be taken is to try to develop a super-ontology or new top-down standard for the types of information that learning systems communicate. While there are roles for such ontologies, top-down ontologies have seldom achieved much support within research or software development communities.

Discussion: Challenges and Research Questions for AIED

Across the prior sections, certain challenges emerged as major challenges for the next few decades of the field. Many of these questions are related to modeling pedagogical

expertise and pedagogical domain knowledge (a longstanding focus of the field), but are reframed due to issues of scaling up (e.g., bigger data sets, varied data, greater content coverage). The following research questions will be increasingly important over the next two decades:

Data Integration:

1. How can learning data from different representations be integrated?
2. How can learning data from different contexts be integrated?

Meta-Adaptation:

1. How and when should data sharing occur between different systems?
2. Who should control handoffs of students between systems, when should they occur, and when should such handoffs be declared to the learner?

Micro-Adaptation:

1. When should researchers optimize for certain types of reward functions?
2. What micro-optimizations produce increased learning?

Pedagogy vs. Domain Expertise in Authoring:

1. What are the key differences between domain knowledge and pedagogical domain knowledge?
2. How can domain pedagogy experts and master authors be identified?

As systems are more commonly combined and interoperating, data science will be a major topic to integrate this data. This topic covers a range of problems, some of which are obvious and some of which more subtle. A longstanding problem for data integration has been mapping between different representations (data structures) and ontologies (semantics). This has already been noted earlier and is already a focus of the AIED community. The general consensus on this problem is that there is no silver bullet, and that either explicit or implicit ontology mapping underlies this problem. Greater integration with Semantic Web and agent-based communication research may be valuable for approaching this problem, but it may remain a longstanding challenge.

With that said, different representations for data do not only mean different data formats or semantic relationships for data in computer systems. They may also mean the qualitatively different kinds of knowledge used by machines versus humans. For example, it is often hard to integrate human expert's mental models with machine learning models. The underlying problem is that machine learning models tend to be statistical inferences from a certain number of observed examples, while expert models tend to be symbolic or subsymbolic representations that were inferred from both an unknown number of observed examples and also other experts' symbolic representations with some assumed level of credibility. A common "solution" to this problem is to collect expert data, then collect training examples for machine learning, and finally throw out the expert data once the machine learning model outperforms the experts on some task. This process describes the de-facto standard for determining knowledge components and their prerequisite

relationships for a set of tasks, as well as other major research topics in AIED. Ideally, future research will need to find models that gracefully integrate both types of data.

A related problem that will become more critical is the issue of integrating data from different contexts. Learning contexts may differ due to the learning tasks (e.g., different AIED systems, different versions), student backgrounds (e.g., culture, goals), time-based issues (e.g., forgetting, changing domain content, stale data), and a variety of other differences. For example, in which cases will data collected in 2015 be meaningful for inferences about students in 2035? To efficiently leverage data, we need to understand how to weigh the relevance of different data sources for making certain inferences. At present, this is often done by what amounts to a hand-picked Boolean filter: data sets are selected for inclusion in an analysis. As data sources expand, this will not be a reasonable option: algorithms will need measures to weight or exclude data based on its relevance or importance for certain types of inferences.

In addition to data integration, data sharing will also be a major issue for meta-adaptation. Meta-adaptation will raise serious issues at both the technical and the societal level. At the technical level, this research needs to find principles or standards for the grain size of information to share. This question boils down to: what kinds of data do different types of learning systems need, and when do they need it? At present, due to the lack of research on the topic, it is not even entirely clear what types of data are useful to improve learning when shared. Estimates of mastery for different knowledge components are probably useful, but beyond that point is quite murky (e.g., Is it useful to know how long a user has been in their most recent session? Would a learner's typing speed be useful to pass along?). This problem is somewhat analogous to the earlier issue of weighting data: how can we estimate the relevance of data in one system to another, and what thresholds for relevance are reasonable for practical applications?

A second major technical issue is that real-life educational systems have many stakeholders with legitimate reasons to control handoffs of students between systems. AIED adaptive systems, teachers, and students themselves all have an opinion on the next learning activity that they think a student should do. Even with a single ITS and a single teacher, it can be difficult to balance the teacher's desire to teach certain topics to a class (e.g., curriculum pacing) against the macro-adaptive loop of an ITS trying to maximize learning for a specific student. Meta-adaptation increases such complexity by an order of magnitude: students might switch between multiple adaptive systems, each with their own macro-adaptive goals, and each student might be either learning independently or as part of a class cohort. There are also questions about when such transitions should be seamless (e.g., like choosing resources in an LMS) or explicit (e.g., like moving between websites using links). Understanding the types of handoffs that both improve learning and are acceptable to educational stakeholders is critical.

At the societal level, managing privacy concerns and data ownership will be another major issue for meta-adaptation. While most academic data is fairly mundane, some types can be quite sensitive. For example, a gaming detector might report that a student is a rampant cheater and an emotion detector might report that a student is quick to anger. Using data to improve learning outcomes must also be balanced with protecting

students, but ethics for data sharing are still being revised to accommodate an increasingly-connected educational world.

Micro-adaptation raises its own serious questions that tie-in strongly to the very nature of educational research: what does it mean to improve education? Reward functions for machine learning have close ties to assessment issues, where significant research is still weighing the implications different measures such as near transfer tasks, far transfer tasks, preparation for later learning, persistence in learning, and motivation for later learning. However, computational reward functions will not accept a paragraph of well-reasoned debate: reward payoffs for different types of tasks will need to be assumed and tuned (e.g., inverse reinforcement learning). Such research will be informed by research on educational assessment, but should also inform how we interpret assessments across education. In the distant future, such questions may even reach the political realm, where legislative debates consider the appropriate reward metrics that educational technology should target.

Micro-adaptation will also open the door for exploring of optimizing different features of the learning experience. This research has strong ties to research on user interfaces, emotion recognition, and dialog-based tutoring, but may eventually explore entirely new data sources and avenues for adaptation. Eventually, such research should outline the holy grail of educational technology: maps of the types of features that are meaningful to adapt to and relative value of different adaptive behaviors for certain learning tasks and contexts.

Next, scaling up AIED will need to push the boundaries of authoring systems for creating content. Currently, work on AIED authoring tends to be atheoretic and lacks standardized processes for assessment and evaluation. Significant work has looked at authoring tools (Murray, Blessing & Ainsworth, 2003), but much more work will still be needed to understand effective methods for authoring different types of AIED learning activities. Moreover, little work has looked at the converse: what makes an effective AIED author? This ties in to the longstanding and controversial question of tutoring expertise: how does one define an expert tutor or teacher? Assessments for identifying domain experts have a long history and are well-established for many domains. However, assessments for domain pedagogy experts (e.g., master teachers, expert tutors) are significantly less clear. This exposes fundamental questions over the dividing lines between domain expertise, domain-independent pedagogical expertise, and domain pedagogy expertise. Understanding the differences between these types of expertise and knowledge for different domains will be an important step for modeling effective instruction in future AIED systems. This is because efficient development of AIED depends on reuse: identifying the space for applying different instructional techniques and domain knowledge is at the heart of that process.

Related to this, to create the best AIED content at a large scale, it will be necessary to develop techniques to identify the best authors: experts in domain pedagogy who are capable to expressing such knowledge using a set of authoring tools. At present, the field lacks validated methodologies to approach this problem. Instead, this tends to be handled as a human resources problem (e.g., hiring based on credentials and interviews) rather than a genuine research challenge for identifying potential experts who know the domain, know how to teach it, and can also master the tools to record such knowledge. In most domains, we have assessments for at most one of these skill sets (i.e., domain expertise) and neither of the other two (i.e., domain pedagogy or

expressing such pedagogy). To create the most effective content, the best authors must be identified. This is because learning gains from AIED adaptation will still be constrained by the underlying content: a useless hint remains a useless hint, no matter when it is presented. As such, methods for discovering “master authors” should be an important area for future research.

In general, moving toward services comes with costs and risks. For example, designing effective distributed services requires different programming expertise: service-oriented and agent-oriented programming means that asynchronous operations become the norm rather than the exception. Second, distributed services increase the risk that a key service may fail or be discontinued. This is particularly relevant for research projects, which seldom have a funding stream for long-term support. Finally, integrating distributed services is a type of collaboration, which requires investment to understand others’ work. While these risks can be mitigated, practices to do so in educational technology are not yet mature.

Closing Remarks

The future for AIED should be a bright one: expansion of learning software into schools will ultimately result in unprecedented diversity and size of user bases. The areas noted in this paper are only the first wave for new AIED opportunities. In time, it will be possible to explore entirely new classes of questions, such as mapping out continuous, multivariate functional relationships between student factors and pedagogical effectiveness of certain behaviors. Systems such as personal learning lockers for data would allow for longitudinal study of learning over time, either in real-time or retrospectively. A major game-changer for future learning research will probably be data ownership and privacy issues: data will exist, but researchers will need to foster best-practices for data sharing, protection, and archiving.

With this wealth of data, researchers will be able to connect learning to other relationships and patterns from less traditional data sources. In 20 years, the range of commonly-available sensor data will be dizzying: geolocation, haptic/acceleration, camera, microphone, thermal imaging, social ties, and even Internet-of-Things devices such as smart thermostats or refrigerators. Moreover, the ecosystem of applications leveraging this data will likewise be more mature: your phone might be able to tell a student not only that their parents left them a voicemail, but that they sounded angry. This event might then be correlated with a recent report card, and the consequences of the interaction might be analyzed. Learning is a central facet of the human experience, cutting across nearly every part of life. To that end, as life-long learning becomes the norm, the relationship between life and learning will become increasingly important. By consuming and being consumed in a distributed and service-oriented world, AIED will be able to play a major role in shaping both education and society.

Acknowledgments The core intuitions for this work have been developed through research supported by the Office of Naval Research (N00014-12-C-0643, W911NF-04-D-0005), the Army Research Lab (W911NF-14-D-0005, W911NF-12-2-0030), and Advanced Distributed Learning (W911QY-14-C-0019). However, the contents and opinions of this paper are the authors’ alone and do not represent those of the sponsoring organizations.

References

- Andersen, E., Gulwani, S., & Popovic, Z. (2013). A trace-based framework for analyzing and synthesizing educational progressions. In W. E. Mackay, S. Brewster, & S. Bødker (Eds.), *SIGCHI Conference on Human Factors in Computing Systems* (pp. 773–782). ACM.
- Arroyo, I., Woolf, B. P., & Beal, C. R. (2006). Addressing cognitive differences and gender during problem solving. *International Journal of Technology, Instruction, Cognition and Learning*, 4, 31–63.
- Banilower, E. R., Smith, P. S., Weiss, I. R., Malzahn, K. A., Campbell, K. M., & Weis, A. M. (2013). *Report of the 2012 national survey of science and mathematics education*. Chapel Hill, NC: Horizon Research, Inc..
- Beck, J. E., & Gong, Y. (2013, January). Wheel-spinning: students who fail to master a skill. In H.C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial Intelligence in Education (AIED) 2013* (pp. 431–440). Berlin: Springer.
- Brusilovsky, P. (1995). Intelligent learning environments for programming: The case for integration and adaptation. In Greer, J. (Ed.), *Artificial Intelligence in Education (AIED) 1995*, (pp. 1–8).
- Cavagnaro, D. R., Myung, J. I., Pitt, M. A., & Kujala, J. V. (2010). Adaptive design optimization: a mutual information-based approach to model discrimination in cognitive science. *Neural Computation*, 22(4), 887–905.
- Chi, M., Jordan, P., & VanLehn, K. (2014). When is tutorial dialogue more effective than step-based tutoring? In S. Trausan-Matu, K. Boyer, M. Crosby, & K. Panourgia (Eds.), *Intelligent Tutoring Systems (ITS) 2014* (pp. 210–219). Springer: Berlin.
- Craig, S. D., Hu, X., Graesser, A. C., Bargagliotti, A. E., Sterbinsky, A., Cheney, K. R., & Okwumabua, T. (2013). The impact of a technology-based mathematics after-school program using ALEKS on student's knowledge and behaviors. *Computers & Education*, 68, 495–504.
- El-Sheikh, E., & Sticklen, J. (2002). Generating intelligent tutoring systems from reusable components and knowledge-based systems. In S. A. Cerri, G. Gouarderes, & F. Paraguacu (Eds.), *Intelligent Tutoring Systems (ITS) 2002* (pp. 199–207). Heidelberg: Springer.
- Falmagne, J. C., Albert, D., Doble, C., Eppstein, D., & Hu, X. (2013). *Knowledge spaces: applications in education*. Springer Science & Business Media.
- Futtersack, M., & Labat, J. M. (1992). QUIZ, a distributed intelligent tutoring system. In *Computer Assisted Learning* (pp. 225–237). Heidelberg: Springer.
- Graesser, A. C. (2011). AutoTutor. In P. M. McCarthy, & C. Boonthum (Eds.), *Applied natural language processing and content analysis: identification, investigation and resolution*. IGI Global: Hershey, PA.
- Heffernan, N. T., Turner, T. E., Lourenco, A. L., Macasek, M. A., Nuzzo-Jones, G., & Koedinger, K. R. (2006). The ASSISTment Builder: Towards an Analysis of Cost Effectiveness of ITS Creation. In G. Sutcliffe & R. Goebel (Eds.), *Florida Artificial Intelligence Research Society (FLAIRS) 2006* (pp. 515–520).
- IMS Global (2015). Learning Tools Interoperability. Retrieved from: <http://www.imsglobal.org/lti/index.html> on Feb 22, 2015.
- Khan Academy (2015). About Khan Academy. Retrieved www.khanacademy.org/about on Feb. 22, 2015.
- Kirschner, P. A., & Erkens, G. (2013). Toward a framework for CSCL research. *Educational Psychologist*, 48(1), 1–8.
- Koedinger, K. R., Baker, R. S., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A data repository for the EDM community: The PSLC DataShop. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S.J.d. Baker (Eds.) *Handbook of educational data mining*, 43–53.
- Liu, Y. E., Mandel, T., Brunskill, E., & Popovic, Z. (2014). Trading Off Scientific Knowledge and User Learning with Multi-Armed Bandits. In *Educational Data Mining (EDM) 2014*. 161–168. Heidelberg: Springer.
- Mizoguchi, R., & Bourdeau, J. (2000). Using ontological engineering to overcome common AI-ED problems. *Journal of Artificial Intelligence and Education*, 11, 107–121.
- Monsalve, S. (2014). A venture capitalist's top 5 predictions for 2015. *Fortune*. Retrieved <http://fortune.com/2014/12/29/a-venture-capitalists-5-predictions-for-2015/> on Feb. 22, 2015.
- Mostow, J., & Beck, J. (2006). Some useful tactics to modify, map and mine data from intelligent tutors. *Natural Language Engineering*, 12(02), 195–208.
- Murray, K., & Silvers, A. (2013). A learning experience. *Journal of Advanced Distributed Learning Technology*, 1(3–4), 1–7.
- Murray, T., Blessing, S., & Ainsworth, S. (2003). *Authoring tools for advanced technology learning environments: toward cost-effective adaptive, interactive and intelligent educational software*. Netherlands: Springer.
- Nye, B. D., Windsor, A., Pavlik, P. I., Olney, A., Hajeer, M., Graesser, A. C., & Hu, X. (2015). Evaluating the Effectiveness of Integrating Natural Language Tutoring into an Existing Adaptive Learning System. In C. Conati, N. Heffernan, A. Mitrovic, & M.F. Verdejo (Eds.), *Artificial Intelligence in Education (AIED) 2015* (pp. 743–747). Cham: Springer.

- R Core Team (2013). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3–900051-07-0
- Ritter, S., Anderson, J. R., Koedinger, K. R., & Corbett, A. (2007). Cognitive tutor: applied research in mathematics education. *Psychonomic Bulletin & Review*, *14*(2), 249–255.
- Roschelle, J., & Kaput, J. (1996). Educational software architecture and systemic impact: the promise of component software. *Journal of Educational Computing Research*, *14*(3), 217–228.
- Shvaiko, P., & Euzenat, J. (2013). Ontology matching: state of the art and future challenges. *Knowledge and Data Engineering, IEEE Transactions on*, *25*(1), 158–176.
- Siemens, G. (2013). Learning analytics: the emergence of a discipline. *American Behavioral Scientist*, 0002764213498851.
- Sottolare, R. A., Goldberg, B. S., Brawner, K. W., & Holden, H. K. (2012). A modular framework to support the authoring and assessment of adaptive computer-based tutoring systems (CBTS). In *Interservice/Industry Training, Simulation, and Education Conference (IITSEC) 2012*.
- Stahl, G. (2010). Guiding group cognition in CSCL. *International Journal of Computer-Supported Collaborative Learning*, *5*(3), 255–258.
- VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, *16*(3), 227–265.
- VanLehn, K. (2015). Regulative Loops, Step Loops and Task Loops. *International Journal of Artificial Intelligence in Education*. Online First.
- Woolf, B. P. (2010). *Building intelligent interactive tutors: student-centered strategies for revolutionizing e-learning*. Morgan Kaufmann.