ACCELERATING LEARNING WITH AN AI TUTOR

Learney AI Ltd[*]
London
United Kingdom

Matthew Phillips
matthew@learney.me

Henry Pulver
henry@learney.me

December 24, 2021

ABSTRACT

Learney is an online AI tutor that accelerates learning for knowledge work. It adapts to the learner by continuously assessing knowledge through short questions. Based on the learner's knowledge, it guides their next steps through questions and explanations, aiming to maximise learning rate towards their goals. It represents knowledge as a directed network composed of concepts as nodes, with edges representing prerequisites (e.g. addition must be understood before multiplication can be). This facilitates finding the most direct path to learn any concept in the network and finding gaps and analogies in knowledge, while providing the learner with a 'map' for contextual understanding and exploring conceptual space. Learney incorporates key findings from cognitive science on learning and memory, such as spatial memory, spaced repetition and flow. The approach is generalisable across domains with a dependency structure - where one concept should be learned before another. The network can expand to new domains through a combination of curation by experts and collective collaboration. The goal is to develop a personalised tool which augments our cognitive abilities, increasing the rate at which knowledge can be integrated into technology.

1 Introduction

The nature of work has changed beyond recognition in recent decades. Knowledge work has risen to be the most valuable form of labour, and in 2019 the world surpassed 1 billion knowledge workers globally (1). In contrast to previous generations who relied on manual labour, knowledge workers' primary source of value-creation is their brain. In the 21st century they will produce more value than the entirety of human history combined. These fields evolve more rapidly than physical labour and involve skills that are more transferable, meaning the modern labour market

[*] https://learney.me/ Web app: https://app.learney.me/

is becoming more fluid. Combined with the increasing value of knowledge work, this means learning is becoming fundamental to our future prosperity.

Software and the internet provide the most scalable means of accelerating learning. But online learning is still in its infancy. Up until now, the priority has been content production and distribution. Universities have published courses online, while creators producing more modular explanations have proliferated. With an ocean of explanatory content now available online, learners drown in choice. As a result, curation is now more valuable than publishing in online learning.

To accelerate learning, curation should be directed by learning progress. But current approaches do not adapt within the learning process, assess either at the end or not at all, or are curated by watchtime. Continuously measuring learning and adjusting the next step provides a feedback loop to guide curation. This can lead to large improvements: studies from cognitive science and pedagogy show a several-fold increase in the rate of learning is possible relative to passive, one-size-fits-all lecture series. We aim to deliver accelerated learning at scale using these insights and measuring and optimising for learning.

This has only recently become possible. Breakthroughs in artificial intelligence in the last decade range from surpassing human-level in the game of Go (2) to revolutionising the understanding of protein folding (3). Similarly, natural language processing has seen extraordinary progress with Transformers (most notably GPT3 (3)) capable of outputting language indistinguishable from human-written text. This enables better understanding of teaching materials, which is key to AI tutoring. Until now, artificial intelligence has primarily been used to automate humans. We believe it should be used to augment human cognition (4). This improves the likelihood of a better future for humanity.

Here we propose a system to accelerate learning. We first review current online learning approaches and the evidence that learning can be accelerated. We then propose a generalisable system that achieves this, personalising to each learner. We discuss the design of community incentives to scale this system. We conclude by speculating about the possibilities it enables.

2 Background

2.1 Limitations of existing online learning

Presently, most online learning is structured as one-size-fits-all courses. Devised in universities, the ‘sage on the stage’ model aims to spread knowledge from one person to many. The majority of the time spent on such courses is in watching passive lectures and the feedback cycle for students tends to be long. At university, it can be months. Online it is shorter, but the exercises can be repeated until guessed correctly rather than meaningfully assessing understanding. And the cycle for adjusting the course is significantly longer, usually annually, with the underlying structure of lecture series and assessments remaining the same. In design terms, this makes the interaction between the object and the user slow and unintuitive, antithetical to well designed systems (5).

Given courses inherit a model intended for in-person learning, it is unsurprising that engagement drops 50% each week through Massive Open Online Courses (MOOCs), achieving only 10% completion rates amongst learners who intend to finish courses (6). The details surrounding delivery in this format matter little - a study of 270,000 students found various behavioural interventions have little effect on outcomes in MOOCs (6). We've all sat through lessons and lectures that bore us or lose us in the first 5 minutes. It's easy to switch off or get distracted when there's no interaction and it's not at your level. Quality of exposition helps, but does not resolve this. Despite recent popularity, cohort-based learning necessitates one-size-fits-all, synchronous learning and increased prices.

The underlying problem of delivering learning at scale can't be solved if everyone receives the same course. Instead, a personalised approach is required. Content should be interactive and tailored to each learner's ability and interests. Further, there is ample reason to believe that personalisation will accelerate learning.

2.2 Accelerating and motivating learning

Research overwhelmingly shows that major improvements in learning can be achieved from simple adjustments. One striking example is spaced repetition. Over a century ago, Ebbinghaus found that rather than focusing exposure to an idea into a narrow space of time, spacing out exposures significantly increases learning - even when the total time spent learning is the same (7; 8; 9). Over one week, spacing exposures leads to memory retention of over 90%, whereas a single exposure declines below 50%. This applies broadly: to learning facts and to learning tasks such as mathematical problems (10). A similarly striking result from Bloom showed that 1-to-1 tutoring achieves a 2 standard deviation increase in test score over classrooms (11). Large increases in learning were also observed using ‘mastery learning’, where students only progress when ready, not according to a prescribed curriculum. In another experiment, test scores were doubled in an undergraduate physics course by increasing the interactivity of teaching, despite being delivered by less qualified instructors (12). It's clear that straightforward changes in the way teaching is delivered can have an enormous effect on learning.

By identifying the underlying psychological reasons for these effects, they can be replicated at scale. Personalisation plays a key role: 1-to-1 tutors can probe the student to test understanding and can tune the level of difficulty accordingly. In psychological terms this helps induce ‘flow’, a state of being fully immersed in an activity (13). To achieve flow, skill and difficulty must be matched, the activity must be interactive, and it must provide rapid feedback. In tutoring, feedback is almost immediate, whereas it's inadequate on online incumbents and slow in educational institutions. This delayed feedback reduces the rate of learning by stifling flow and because the time gap between action and outcome leads to the ‘credit assignment’ problem [2].

Motivation also plays a critical role. In school, impending mandatory exams and teachers serve as extrinsic motivators to push through unengaging material and delivery. Online these aren't present, so learners drop off MOOCs, leading to poor retention figures. But online learning is still structured around extrinsic reward. This was studied compellingly by putting children in a room alone with a notepad and pencils, and comparing promising them a reward if they draw with no reward offered. The children spent over twice as much time drawing with no reward offered (15)). We see this alternative play out successfully in Montessori schools, where intrinsic motivation is encouraged by giving learners autonomy. Students learn what interests them, at their own pace, not through a predetermined curriculum. The result is substantial increases in maths, literacy, and a range of soft skills (16). The same principle can be applied to

[2] The credit assignment problem as formulated originally in AI systems by Minsky (14), where ambiguity over the actions that led to the outcome accumulates with time. Interactivity and feedback time are therefore vital for learning

asynchronous online learning. Intrinsic motivation can be cultivated by enabling autonomy, matching difficulty to skill on interactive content, and demonstrating mastery through clearly visualised progress.

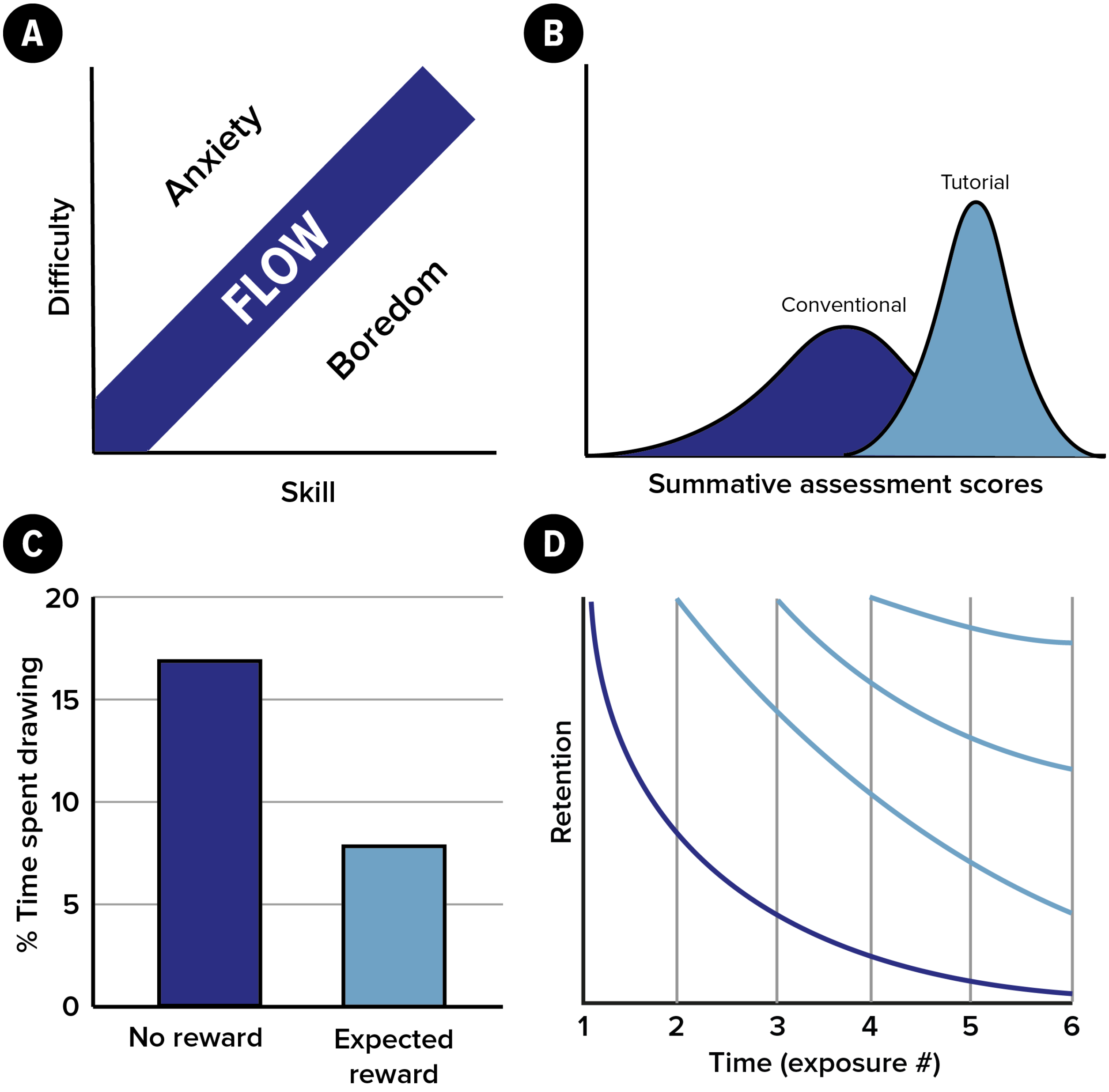

Figure 1: **Cognitive science research shows we can accelerate learning.**
(A) The concept of 'flow', as outlined by Csikszentmihalyi. To induce flow, the difficulty of the task must be matched to the skill of the person undertaking it. **(B)** Bloom 1982's finding that 1-to-1 tutorials achieve an improvement in test results 2 standard deviations above conventional teaching approaches. **(C)** Lepper and colleagues finding that when given the chance to draw, children encouraged through a reward spent less time drawing. This shows that in curiosity driven pursuits, intrinsic motivators are more powerful than extrinsic motivators, though the latter is currently the dominant motivational approach used. **(D)** Ebbinghaus's finding that spacing exposure to concepts in time leads to greater knowledge retention.

Previous successful efforts have been made in building a digital tutor, but not at scale. In 2009, DARPA funded the production of a digital tutor to train novices in IT skills. Despite the program only running for 16 weeks, compared to the 35 week in person training program, students who were taught through the tutor had more than 2 standard deviations better outcomes - outperforming even Bloom's 1-to-1 tutoring study. This is an astonishing result, suggesting the key component of a tutor is the identification of misconceptions, not motivation (17; 18). The task ahead is to design a system that can achieve this at scale.

Our overall contention is that combining these straightforward changes to how online learning is delivered will make a dramatic improvement to learning outcomes. This requires personalisation - a system that adapts to the learner.

3 Learney: Personalising learning

3.1 Adaptive learning system design

This is a sequential decision-making problem, so can be formalised as a Markov Decision Process (MDP). This comprises an agent interacting with an environment which receives feedback (reward) based on the actions it takes - a class of solutions called Reinforcement Learning (19). In the case of tutoring, the tutoring system is the agent, the state comprises the knowledge, ability and goal of the student, while possible actions are items of learning material that can be presented. Our proposed reward signal is progress towards the learner's goals, discussed further in section 3.2. However, mastering automated tutoring requires overcoming several long-standing challenges in solving MDPs. These include imperfect information (the state isn't directly observable), a massive space of possible actions and a delayed reward signal - learning progress can only be inferred after-the-fact (20). [3]

To simplify this problem, Learney structures knowledge as a dependency graph. The nodes represent concepts and edges are dependencies - e.g. to learn about squaring a number, one must first understand multiplication. We think of this as a 'map' that can be navigated. First, the learner sets their goal - understanding a set of nodes to a chosen depth. This can be learning any concept (e.g. matrix multiplication), application (e.g. surfaces in 3D graphics) or entire field (e.g. linear algebra) to the depth they would like. The map defines the path of concepts to reach the user's goals, reducing the search space for content to the boundary of the learner's abilities on that path.

To make effective recommendations, the student's knowledge and ability must be modelled. These can't be measured directly, but can be inferred from user interactions. Metrics on passive learning material, such as watch-time on videos, do not give enough information to effectively infer understanding. A fully-watched video could signify a learner understood easily, struggled but persevered, or left for a coffee. In contrast, a learner's answers to short questions provides sufficient information for effective inference of their knowledge. Short questions are most effective as learners receive feedback faster, can more precisely identify misunderstandings, and they give more data points for the system to learn from. This, coupled with the map of prerequisites, enables identification of misconceptions [4] and gaps in knowledge.

[3] There is an intriguing dynamic whereby the actions taken must both probe the learner's knowledge, as well as teach the learner. Given the short questions serve dual purposes - both assessing learners and giving them practice - both can be achieved in parallel.

[4] Misconceptions can be identified through the type of incorrect answer given - so called 'diagnostic questions'.

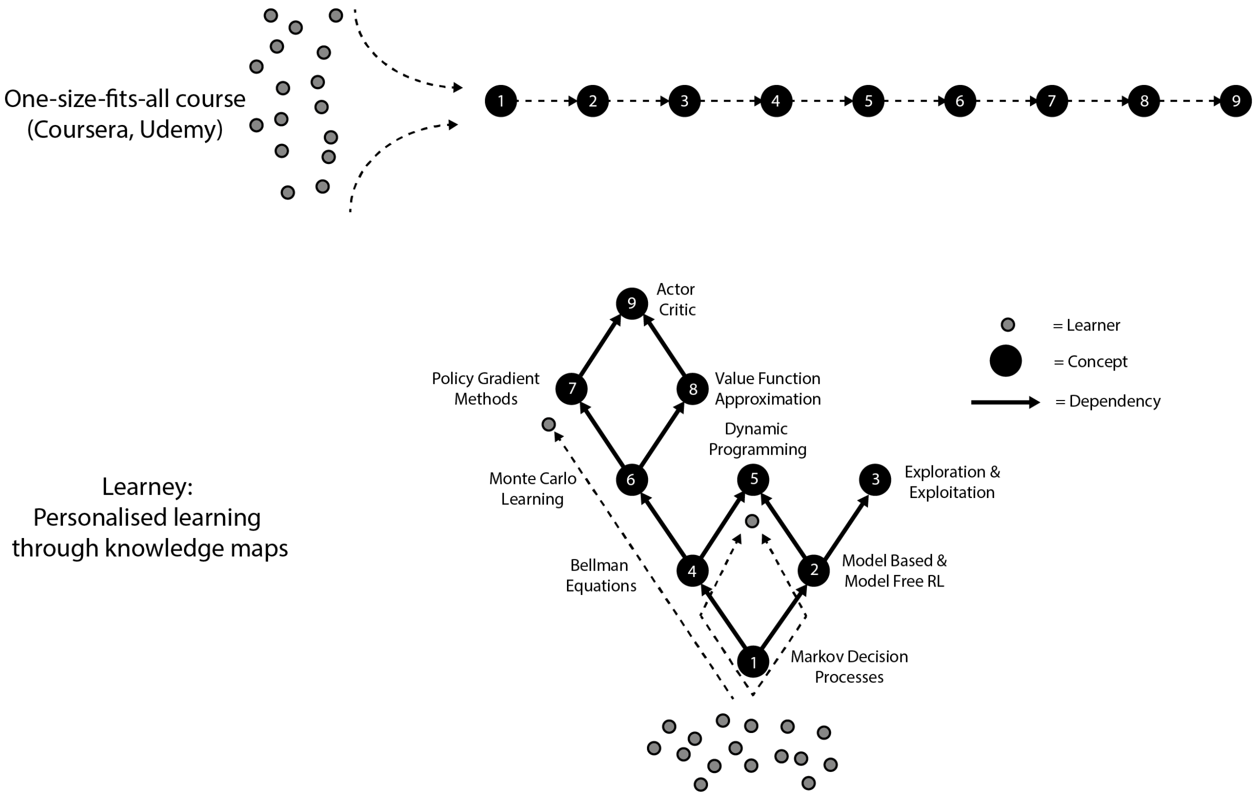

Figure 2: **Computing paths through a knowledge graph saves learners time.**
(A) The current 'one-size-fits-all' model means all learners take the same path, regardless of their prior knowledge or goal. **(B)** Knowledge graphs with learning dependencies between concepts enable paths to be computed for each learner, saving them time by cutting out unnecessary concepts.

Modelling knowledge as dynamic can account for both learning and forgetting. This enables Ebbinghaus-like forgetting curves after exposure to each idea to predict forgetting, enabling optimal spacing of concepts through time. Such curves, when learned empirically for each user, show a marked improvement over the theoretical approaches (21). The ability to model knowledge has increased in other ways, with deep learning approaches being applied to 'knowledge tracing' and the selection of appropriate explanations (22). We therefore anticipate achieving further learning gains beyond those from applying insights from cognitive science by optimising for learning.

The Learney system taken in its totality comprises:

- A set of linked pedagogical knowledge maps
- Content associated with each node on the map, including both explanations and question sets
- An algorithm for selecting the next item of content to be shown
- A set of incentive structures for community and expert contribution

The students' perspective will be to arrive at Learney and enter what they would like to learn. They will see the overview map of the field, and be given an appropriate initial test to tune the difficulty level. The adaptive learning system will then select a path through the dependencies in knowledge to find them the fastest way to learn their goal. Highly modular content - both explanations and questions - will then be shown in the appropriate order, recomputing after each step.

3.2 Measuring learning

Optimization requires an objective. To maximally accelerate learning, this should be the learning progress achieved by a learner thanks to an item of content towards their goal. Given continuous assessment through questions, this progress can be measured. We introduce a method to measure this learning progress achieved. We estimate the time taken to reach the learner's goal before and after a piece of content. This is a deterministic function of the learner's knowledge, goal, ability and learning habits [5], learned from data. After a learning session, we estimate **time-to-goal** before and after each piece of material shown and find the difference, ΔTTG. Subtracting the time taken on the material from ΔTTG gives the difference between the estimated time to perform the learning and actual time spent learning. ΔTTG can then be used as a 'reward signal' to be optimised by the adaptive learning system, and to replace watchtime as an incentive for creators, moderators and community contributors to be rewarded.

3.3 Structuring knowledge for personalised learning

By mapping knowledge, we can better understand and identify goals and misunderstandings. The visual representation has other benefits - contextualising concepts for learners, enabling more structured exploration and saving time over predetermined curricula by finding the most direct path to learn a concept. Structuring knowledge as a map enables learners to see how a concept fits into a broader picture and how it relates to ideas they already understand. Spatial cognition is particularly powerful due to evolutionary history. Selective pressure to occupy advantageous positions favoured organisms that model the spatial environment around them (23; 24). Recent evidence suggests these spatial representations adapted to form a general cognitive system where different modalities can be modelled in spatial terms (25; 26). The results of this trick of evolution are seen in modern society, including by world memory champions using the 'method of loci', in which familiar locations are used to remember long sequences, including the order of entire decks of cards in under a minute (27). The presentation of concepts as a map is an intuitive interface for users to learn from.

3.4 Recent advances enabling an AI tutor

It has only become possible to build a general purpose system of this kind in recent years due to technological developments. In the last 5 years, machine learning approaches have seen breakthroughs in performance on sequential decision-making problems - surpassing human ability in games such as Go, Chess, Starcraft and Poker (2; 28; 3). These approaches formulate the problems as MDPs. They are already applied by consumer-facing companies in recommendation systems, such as Netflix and TikTok. Most excitingly, they show successfully overcoming many of the long-standing problems with notable progress in dealing with large action spaces and imperfect information (28).

Advances have also been made in the ability to produce natural language. Large language models such as GPT3 enable language analysis, text summarisation and text synthesis. This brings within reach effective question generation on

[5] Learning habits may have the largest effect on progress. Highly infrequent practice leads to forgetting, meaning more time must be spent revisiting concepts, slowing progress.

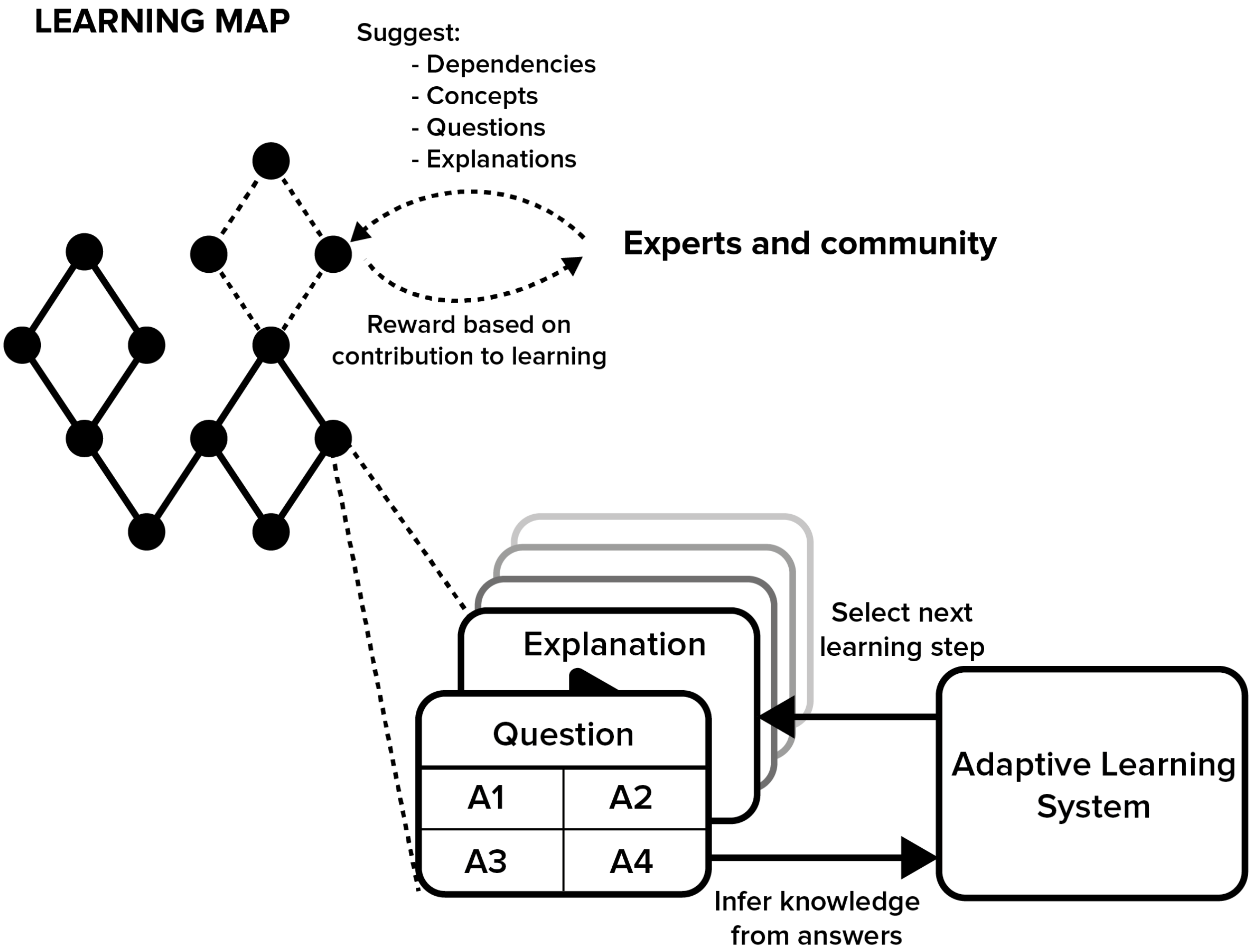

Figure 3: **Overview diagram of the Learney system.**
A knowledge graph built by experts and a community enables computing personalised paths. Rewards are then attributed to contributors based on their contribution to learning. Learning on the path happens through the presentation of short questions and explanations, which are selected through an adaptive learning system.

given concepts. This enables large-scale automated question generation, with the potential for generating questions and explanations unique to the specific learner. This would provide a fully personalised experience and is a promising long-term approach for Learney.

3.5 Collaborating to grow an adaptive learning platform

To build a knowledge map that covers many subjects while maintaining pedagogical integrity, some form of intelligent oversight is necessary. We propose a combination of experts creating pedagogical maps and a Wikipedia-like crowd-sourcing of the mapping process. Incentivising effective mapping is crucial to achieving this.

We propose contributors be rewarded according to their contribution to learning. Whereas other platforms prioritize watch-time due to monetising through adverts or time-to-purchase due to monetising through certificates, our priority is to provide the most efficient way to learn. Through measuring learning we can align incentives across the platform towards this aim either through reputation systems or a cryptographic token.

We propose three types of contribution:

- Content: including question writing and suggesting and creating explanatory content
- Pedagogical mapping: including by dedicated community members, subject experts, and creators
- Moderation: by dedicated community members

Each of these can be rewarded according to their contribution to ΔTTG. By doing so, we provide a means of assessing content and community contribution. The precise mechanics of the reputation system underlying this reward for contributors will be adjusted according to community needs, but we take platforms such as StackOverflow, Wikipedia and Reddit as examples.

A single, unified map enables useful extensions. We can plan paths through the map to find the fastest route. Moreover, we can transfer progress across subjects. For example, Quantum Computing and Machine Learning both require probability: a learner who changes goals from one to the other shouldn't cover this twice from scratch. Each of these saves learners valuable time while learning.

3.6 Roadmap

Now. So far we have tested each component of this system in isolation. We have built a prototype map of the mathematical prerequisites of machine learning. We have run initial trials of short questions delivered through a chatbot, which users like so much they are writing questions to keep access to it. Further, we have developed a basic map editor, enabling users to build their own maps and have our first contributors.

Learney v1.0: Full product. Our next stage is to integrate these features together. This will allow us to measure learning progress. Once we've built low-friction means of contributing, we can tie contributions to this measurement: with a reward system that is either purely reputational or uses a cryptographic token. We will also develop a mobile application that enables users to learn on-the-go. Lastly, we will begin using machine learning to recommend questions and explanations, aiming to maximise the measured learning progress.

End goal. In the long term, we aim to scale to all science, technology, mathematical and related fields. The only limitation on fields we can cover is a clear pedagogical dependency structure. Learney benefits from a 'fly-wheel effect', where growth in users improves both the recommendation system and the number of concepts covered, leading to more users.

4 The future of learning

The impact of accelerating learning would be profound. In the modern knowledge-based economy, learning underpins human advance. But we currently spend little time improving how we learn. The first order effect of doing so would be to enable people to more quickly acquire and apply a greater breadth and depth of knowledge. We conclude here by speculating about second order effects enabled by widespread adoption.

For scientists, entrepreneurs and engineers, filtering and learning from the deluge of new knowledge is one of their greatest challenges. For these knowledge workers, learning is essential to solve new problems and improve upon existing technology. As these workers have the potential to produce solutions that scale to all of us, increasing the rate at which they can achieve this will lead to greater prosperity for all.

Our system provides a means of verifying knowledge. By testing with questions associated with particular concepts, we can verify a student's knowledge base at much higher resolution than certification through current assessments or exams can. For example, we can imagine sending a verified knowledge graph alongside a CV for job applications. This is vital in fields where knowledge is developing so fast that the content covered by a credential is out of date by the time it is issued or obsolete soon after.

Mapping and verifying knowledge can also be used within organisations to understand and deploy their knowledge and skills effectively. When selecting teams to work on projects, an in-depth understanding of their expertise lets companies find skills gaps, preemptively train employees, and identify biases attached to specific ways of thinking. And by verifying learning, companies can much better assess the return on investment from learning.

The soaring value of knowledge work increases the importance of learning. While learning is no longer beholden to institutional gatekeeping, the predominant approaches to verifying knowledge still follow their model. We are building the system to accelerate the rate at which people learn and push forward human progress.

References

[1] Forbes Technology Council and Sébastien Ricard, “The Year Of The Knowledge Worker,” *Forbes*, 12 2020.

[2] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, “Mastering the game of Go without human knowledge,” *Nature*, vol. 550, 10 2017.

[3] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, “Language Models are Few-Shot Learners,” 5 2020.

[4] J. Heer, “Agency plus automation: Designing artificial intelligence into interactive systems,” *Proceedings of the National Academy of Sciences*, vol. 116, 2 2019.

[5] D. Norman, *The Design of Everyday Things*. New York: HarperCollins, 1973.

[6] R. F. Kizilcec, J. Reich, M. Yeomans, C. Dann, E. Brunskill, G. Lopez, S. Turkay, J. J. Williams, and D. Tingley, “Scaling up behavioral science interventions in online education,” *PNAS*, vol. 117, pp. 14900–14905, 6 2020.

[7] H. Ebbinghaus, “Memory: A Contribution to Experimental Psychology,” *Annals of Neurosciences*, vol. 20, 10 2013.

[8] R. F. Hopkins, K. B. Lyle, J. L. Hieb, and P. A. Ralston, “Spaced Retrieval Practice Increases College Students’ Short- and Long-Term Retention of Mathematics Knowledge,” *Educational Psychology Review*, vol. 28, pp. 853–873, 12 2016.

[9] H. F. Spitzer, “Studies in Retention,” *The Journal of Educational Psychology Studies in Retention*, vol. 30, no. 9, pp. 641–656, 1939.

[10] D. Rohrer and K. Taylor, “The shuffling of mathematics problems improves learning,” *Instructional Science*, vol. 35, pp. 481–498, 11 2007.

[11] B. S. Bloom, “The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring,” *Educational Researcher*, pp. 4–16, 1982.

[12] L. Deslauriers, E. Schelew, and C. Wieman, “Improved Learning in a Large-Enrollment Physics Class,” *Science*, vol. 332, 5 2011.

[13] M. Csikszentmihalyi, “If We Are So Rich, Why Aren’t We Happy?,” tech. rep., 1999.

[14] M. Minsky, “Steps toward Artificial Intelligence,” *Proceedings of the IRE*, vol. 49, 1 1961.

[15] M. R. Lepper, D. Greene, and R. E. Nisbett, “Undermining children’s intrinsic interest with extrinsic reward: A test of the "overjustification" hypothesis.,” *Journal of Personality and Social Psychology*, vol. 28, no. 1, 1973.

[16] A. Lillard and N. Else-Quest, “Evaluating Montessori Education,” *Science*, vol. 313, 9 2006.

[17] J. A. Kulik and J. D. Fletcher, “Effectiveness of Intelligent Tutoring Systems,” *Review of Educational Research*, vol. 86, 3 2016.

[18] J. D. Fletcher, “The value of digital tutoring and accelerated expertise for military veterans,” *Educational Technology Research and Development*, vol. 65, 6 2017.

[19] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, 2018.

[20] G. E. Monahan, "State of the Art—A Survey of Partially Observable Markov Decision Processes: Theory, Models, and Algorithms," *Management Science*, vol. 28, 1 1982.

[21] B. Settles and B. Meeder, "A Trainable Spaced Repetition Model for Language Learning," *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pp. 1848–1858, 8 2016.

[22] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. Guibas, J. Sohl-Dickstein, S. University, and K. Academy, "Deep Knowledge Tracing," tech. rep., 2015.

[23] K. J. Jeffery and C. Rovelli, "Transitions in Brain Evolution: Space, Time and Entropy," *Trends in Neurosciences*, vol. 43, 7 2020.

[24] J. O'Keefe and L. Nadel, *The Hippocampus as a Cognitive Map*. Oxford: Clarendon Press, 1978.

[25] D. Aronov, R. Nevers, and D. W. Tank, "Mapping of a non-spatial dimension by the hippocampal-entorhinal circuit," *Nature*, vol. 543, pp. 719–722, 3 2017.

[26] J. L. Bellmund, P. Gärdenfors, E. I. Moser, and C. F. Doeller, "Navigating cognition: Spatial codes for human thinking," 11 2018.

[27] K. A. Ericsson, W. G. Chase, and S. Faloon, "Acquisition of a Memory Skill," *Science*, vol. 208, 6 1980.

[28] O. Vinyals, I. Babuschkin, W. M. Czarnecki, M. Mathieu, A. Dudzik, J. Chung, D. H. Choi, R. Powell, T. Ewalds, P. Georgiev, J. Oh, D. Horgan, M. Kroiss, I. Danihelka, A. Huang, L. Sifre, T. Cai, J. P. Agapiou, M. Jaderberg, A. S. Vezhnevets, R. Leblond, T. Pohlen, V. Dalibard, D. Budden, Y. Sulsky, J. Molloy, T. L. Paine, C. Gulcehre, Z. Wang, T. Pfaff, Y. Wu, R. Ring, D. Yogatama, D. Wünsch, K. McKinney, O. Smith, T. Schaul, T. Lillicrap, K. Kavukcuoglu, D. Hassabis, C. Apps, and D. Silver, "Grandmaster level in StarCraft II using multi-agent reinforcement learning," *Nature*, vol. 575, 11 2019.