AI's Recent History

The Skinny on AI for Education, September 2023

Headlines

- **The Main Read: Getting to grips with AI is no longer an option for those involved in education and training, it is essential**
- **The Cutting Edge: AI and the future of education**
- **News: Microsoft's smart AI, uses of ChatGPT, and Stable Diffusion**
- **Regulation, Ethics & Safety: Digital platform regulation, public working groups on AI, and OpenAI's Head of Trust and Safety departs**

The Main Read: Getting to grips with AI is no longer an option for those involved in education and training, it is essential

The way technology innovation plays out once it's released 'into the wild' that is the real world is not easy to predict and requires an understanding of what the technology can achieve and an understanding of the context in which the technology is being applied. One only has to look at the way that the technology that has been used to reduce staff costs in shops has led to a certain lawlessness when it comes to shoplifting to see what I mean. There are no longer enough staff to watch who is taking what and self-checkout tills largely just leave customers to police themselves. The situation with AI will certainly not be any more straightforward and likely far more complicated. So, when it comes to education, we really must ensure that the people who understand the context also understand that technology and a part of the conversation about how that technology is applied. *There is little doubt that what is happening now with AI* is transforming our world and therefore for anyone involved with education or training, *engaging with AI is no longer an option, it is essential.*

What have I been reading?

The extent to which AI is changing business and the workplace are a constant subject of debate. A recent report from McKinsey proposes widescale industrial benefit from Generative AI, although at the moment customer service, marketing/sales, software engineering, and R&D are the key areas of activity. So, it might be easy to think that education and training can sit back and wait for 'their turn' to come. However, the report suggests that Generative AI can automate 60-70% of employees' work time by understanding natural language, and that this affects higher wage knowledge work in particular. The report predicts that more than half of today's work activities will be automated by 2030-2060, a decade earlier than previous estimates and stresses that **realising the benefits of AI will require that we determine the new skills that are needed and rethink the training that is provided.**

Reports about automation impacting on jobs are not new, but many of the calculations and predictions are being updated to take account of the release of Generative AI. In addition to the report from McKinsey, there are several others that look at the impact of new AI tools on the workplace, including a report from researchers at OpenAI and UPenn which also highlighted the extensive exposure of white-collar jobs to automation driven by AI. And the Partnership on AI, who released a report that recognises the role that AI developers play when it comes to the impact of AI on the workplace. Their decisions and choices need to be made wisely and the PAI report crafts a set of Guidelines for AI and Shared Prosperity, which aims to provide the conceptual tools to steer us towards improved job access and improved job quality and is accompanied by a high-level Job Impact Assessment tool, but this alone will not help drive productivity and career fulfilment for young people.

The significant implications of what is happening with AI for education and training is crystal clear. **Without question, education and training are more important now than ever before and our systems must now develop the capacity to expand our human intelligence to ever more sophisticated levels, way beyond what AI can achieve.** I am certainly not alone in thinking about richer and deeper Human Intelligence in these time of AI enthusiasm, an interesting article from the Brookings institute celebrates the importance and value of curiosity and creativity. This is not the time to dumb down our human intelligence because we've developed smart AI that can 'ace' most assessments

and that is full of confidence in its ability to do most of what we ask. Now is the time to rethink what we mean by the word intelligence and to focus on developing the aspects of human intelligence that cannot be automated.

There is also an important role for technical skills and I enjoyed reading 'In praise of the 'techies' who make companies more productive. The authors celebrate the way that technical skills and the ability to implement new technology matter. They refer to research that illustrates that "techies" who install, maintain and optimize technology boost productivity across companies and sectors and suggests that Britain needs more of these tech-savvy workers to benefit from AI and automation. There have been many calls for more and different approaches to technical education (see for example). But with the current wave of AI, makes it more urgent that we help people gain both a non-technical AI literacy and for some people, a detailed technical understanding of AI. And yet, how we help people gain the AI skills and knowledge they need when the technology is evolving so quickly, is not obvious and requires flexibility as well as speed: not a combination that is straightforward to deliver.

The Cutting Edge: AI & the future of education

Duolingo, known for its popular language learning app, appears to be developing a new app focused on learning music concepts like reading sheet music and playing instruments. This shows how education companies are expanding into new subjects as technology allows for customized, self-paced learning beyond the classroom. Music education could benefit from AI features like real-time feedback. One must wonder though if there might be an element of business concern about the future of language learning when translation is becoming increasingly accurate for both written and spoken language. Let's hope that our interest in learning languages persists, because there is something beautiful about learning how people communicate across different cultures.

News: Microsoft's smart AI, uses of ChatGPT and Stable Diffusion

In other AI news, Bill Gates recently chatted with Khan Academy founder Sal Khan about how tools like Khan Academy's KhanAmigo (Khan's new AI tutor) can supplement teachers, not replace them. It is an interesting discussion and you can read the transcript or listen to the pod. Sal explains how AI assistants can help engage students, facilitate peer interactions, and provide personalized explanations to fill knowledge gaps. This resonated with Bill, who sees major potential for AI to close achievement divides if thoughtfully implemented. Transforming education requires motivated teachers who inspire curiosity. AI is simply a tool to help teachers maximize their impact and there are some interesting reflections about how memorable and influential good teachers are.

Microsoft and other tech giants are racing to integrate smart AI features into their products. For Microsoft, this means infusing AI throughout Office apps to provide automated assistance. Outlook could summarize emails and suggest responses. Word could rewrite reports in different styles and lengths. PowerPoint could create presentations by analysing raw data and notes. The potential to enhance office productivity with AI is immense.
But AI may also emerge as a whole new computing platform beyond current devices and apps. Right now, tech firms compete by building better hardware and software capabilities. But mastering cutting-edge AI models could soon become a prerequisite for future dominance. AI engines like GPT-4 could disrupt the industry like Windows and iOS did before. Rather than being an add-on, foundational AI could define the next era of technology. Control of the top AI platforms may determine who rules big tech.

An interesting and perhaps unpredicted use for ChatGPT: Iowa educators are using ChatGPT to identify books to ban from school libraries to comply with new Republican laws restricting "inappropriate" content, leading to 19 books being removed so far in Mason City. The bans attempt to follow laws requiring school libraries be "age appropriate" and not contain sexual descriptions or depictions.

More vibrant image generation available here: AI startup Stability AI has launched Stable Diffusion XL 1.0, an upgraded text-to-image model that produces more vibrant, accurate images with better lighting and contrast compared to the previous version. The new model can generate full 1-megapixel

images faster and is improved at generating complex images from text prompts and producing legible text.

Regulation, Ethics and Safety: Digital platform regulation, public working groups on AI, and OpenAI's Head of Trust and Safety departs

There is always a lot of activity in this space and many challenges to be tackled. Here are some of the developments that I found interesting:

Policy researchers and advocates have called on legislators to establish a new agency to regulate digital platforms: an interesting article from the Brooking Institute. And there is more…

The U.S. Commerce Department is launching a new public working group on AI to build on its AI Risk Management Framework and address opportunities and challenges of generative AI. The group will help National Institute of Standards and Technology (NIST)develop guidance for managing risks of generative AI like text, image and video generation, and support testing and measurement. In the long-term, the group will explore using generative AI to address challenges in health, environment and climate. This comes as the White House meets with AI experts and receives recommendations from its advisory committee on focusing AI efforts.

OpenAI's head of trust and safety Dave Willner is leaving the company due to job pressures affecting his family life, though he will continue advising through the end of the year. His departure comes amid rising concerns about the potential negative impacts of AI that trust and safety teams at tech companies are trying to address.

It seems that there is no easy way to control what GenAI produces… Despite developers' best efforts. researchers showed that adding certain prompts can get ChatGPT and other AI chatbots to spit out undesirable content like hate speech, despite defences put in place to prevent this. The findings indicate the tendency for advanced chatbots to go astray stems from a fundamental weakness, complicating efforts to safely deploy the most capable AI systems.

There is no easy answer to detect plagiarism with GenAI: OpenAI shut down its text classifier tool that aimed to detect AI-generated writing, after admitting it had low accuracy and could falsely flag human writing as AI. OpenAI says it will now focus on developing more effective techniques for detecting AI-generated content across text, audio and visual media.

There are continuing concerns about ever more sophisticated Deepfakes that add to concerns raised by the amount of investment being made into companies that develop these technologies, something I highlighted in The Skinny Issue 2: Deepfake technology can create highly realistic fake videos of people, going beyond traditional editing to make digital puppets and blurring the line between real and fake. With few regulations so far, experts warn deepfakes could enable widespread disinformation, as shown by a recent fake video of Ukraine's president announcing surrender.

Regulating AI is complicated by the connections each AI technology has to yet more technologies, such as plugins. While plugins expand ChatGPT's capabilities, their lack of validation means OpenAI's model can't trust what plugins return, creating vulnerabilities to attacks: ChatGPT impresses people with its text generation abilities, but is limited by its fixed dataset from 2021 and inability to access new web data. Plugins allow developers to add functionality to the paid GPT-4 version, like flight booking and document analysis, but security researchers warn they could expose data or allow remote code execution.

A victory for legislators targeting AI bias: China-based tutoring company iTutorGroup agreed to settle a lawsuit by the U.S. Equal Employment Opportunity Commission (EEOC) alleging its hiring software used AI to illegally discriminate against older job applicants. This 2022 case was the EEOC's first lawsuit involving a company's use of AI in hiring decisions. The EEOC warns it will target employers misusing AI for discrimination, as surveys show over 85% of large U.S. companies now use AI in employment in some way. Experts expect more lawsuits as workers accuse employers' AI systems of bias against characteristics like race, disability and age.