Data Analytics – *Lecture Notes*

**In general,** data analytics is defined as the process of examining data sets in order to draw conclusions about the information they contain. Accountants already work with data and data analytic outputs to assess business performance within the framework of accounting principles and methods and to identify trends and unusual items requiring further investigation.

The AUD Blueprint changes require CPA candidates to demonstrate their ability to use data analytic outputs when “identifying and assessing risks of misstatement due to error or fraud and developing appropriate engagement procedures” as well as when “performing engagement procedures and concluding on the sufficiency and appropriateness of evidence obtained.”

Digitization of the enterprise has altered not only the speed at which change happens but also created demand for faster ROI.

https://www.cfo.com/it-value/2019/04/advanced-technologies-are-taking-over-finance/

*Video*:

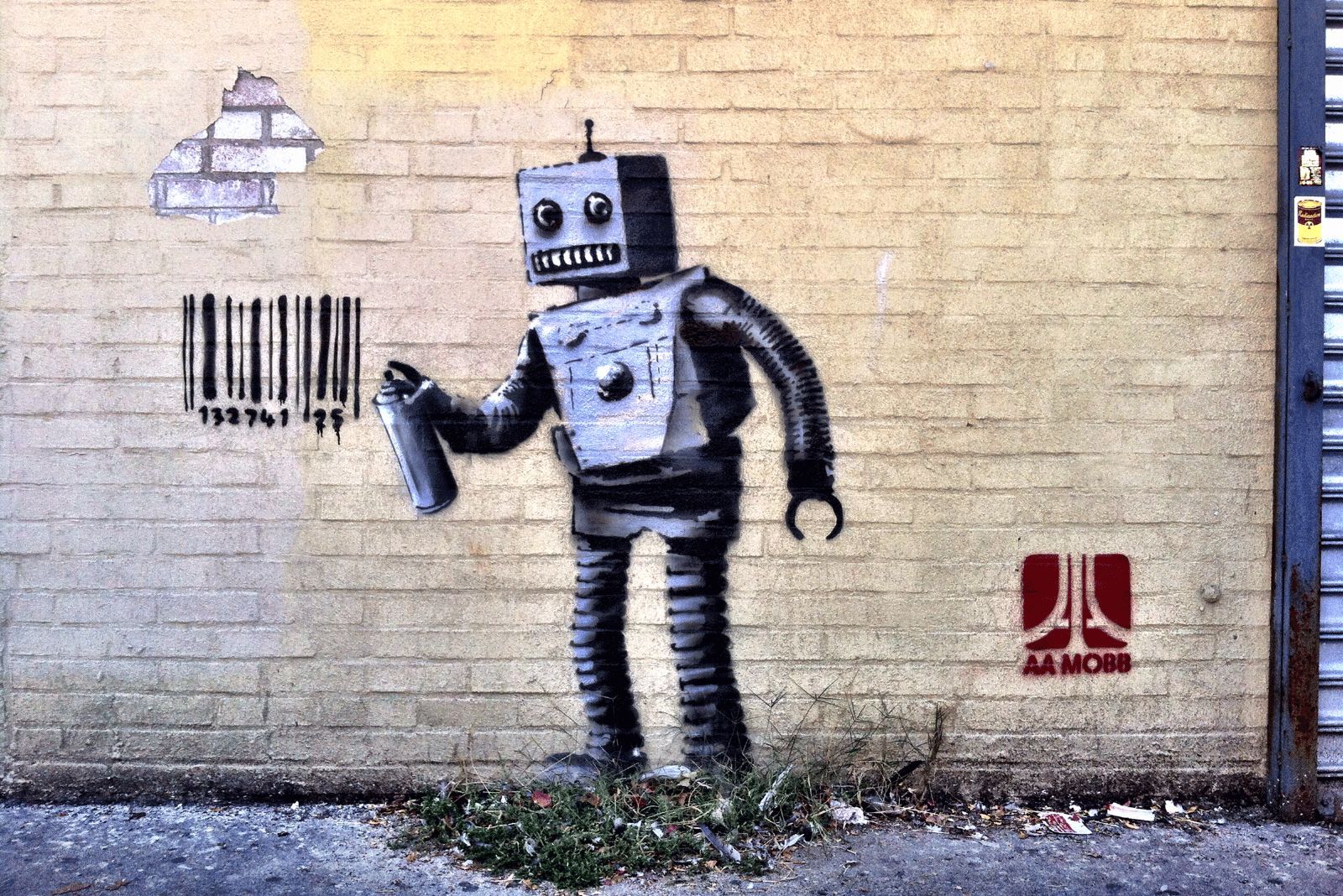
With the advances in technology, the audit is poised for rapid change – beyond the traditional financial statement audit. What must auditors do to ensure they are prepared to meet the demands of audit in the future?

https://www.youtube.com/watch?v=6qPZJfe5jXc

Data analytics is enhanced by robotic learning and computers programing themselves, called self-coding AI. The most recent is DeepCoder, a machine learning system that can write its own code. It does this using a technique called program synthesis. Essentially, it creates new programs by combining existing lines of code taken from other software, which is what human coders do. With a particular output in mind, DeepCoder was able to determine which lines or pieces of code would be particularly useful.

“The approach is to train a neural network to predict properties of the program that generated the outputs from the inputs.

https://futurism.com/4-our-computers-are-learning-how-to-code-themselves

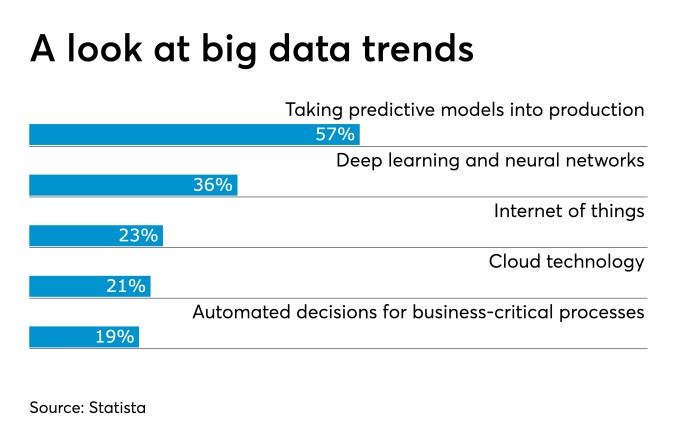


**AI & Robotics**

Creating value with human-like robots. For more than two decades, Hanson Robotics has been building the world’s most human-like robots. Endowed with rich personality and holistic cognitive AI, our robots are able to engage emotionally and deeply with people.  They can maintain eye contact, recognize faces, understand speech, hold natural conversations, and learn and develop through experience.

**david henson ai**

***video*: https://www.youtube.com/watch?v=1y3XdwTa1cA**



Enterprise- and consumer-technology companies attach the AI label to a wide range of products — among them chatbots, facial recognition, and even smartphone games that “predict” how you will look when you grow old. So, what is AI, and what would it allow users such as CFOs and their teams to achieve?

While AI is a broad term, it all comes down to data and getting a learning engine to do a better job, according to Eric Thain, co-chair and president at the Artificial Intelligence Society of Hong Kong and general manager of brand development at HK Express, a Hong Kong-headquartered airline.

As machines learn by trial-and-error and/or pattern recognition, data is the fuel for their learning process, Thain said in a phone interview. “The more the data, the better the learning engine works,” he noted.

In addition, machine learning is the natural progression of data analytics when it comes to AI, he said.

**https://www.fm-magazine.com/news/2019/oct/how-to-start-ai-deployment-201921630.html?utm\_source=mnl:cpald&utm\_medium=email&utm\_campaign=10Oct2019**

<https://www.theepochtimes.com/google-achieves-quantum-supremacy-that-will-soon-break-all-encryptions_3096063.html>

**Machine learning** is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention. https://www.sas.com/en\_us/insights/analytics/machine-learning.html

Data science, analytics, and machine learning are growing at an astronomical rate and companies are now looking for professionals who can sift through the goldmine of data and help them drive swift business decisions efficiently.

3 million new jobs by the end of next year…

https://www.simplilearn.com/data-science-vs-data-analytics-vs-machine-learning-article

***TIME as a data point…***

***The probabilistic framework***, which describes how to represent and manipulate uncertainty about models and predictions, has a central role in scientific data analysis, machine learning, robotics, cognitive science and artificial intelligence.

The key idea behind the probabilistic framework to machine learning is that learning can be thought of as inferring plausible models to explain observed data. A machine can use such models to make predictions about future data, and take decisions that are rational given these predictions. Uncertainty plays a fundamental part in all of this.

<https://www.nature.com/articles/nature14541>

Data, analytics and machine learning are the foundation for AI (artificial intelligence). The challenge with data is the variety across locations (cloud, on-prem, private cloud), types (structured, unstructured), and platforms (operational database, data warehouse, hadoop, fast data platforms, etc). Once we deal with our data management, we are able to move on to analytics, which lets us extract insight from our data. Predictive analytics leads us to machine learning. Once we develop enough machine learning models, we are beginning to reach AI.

YouTube Video: https://www.youtube.com/watch?v=Db24-JRN3ls

Data Analytics on the CPA Exam

https://www.rogercpareview.com/blog/learn-how-data-analytics-is-coming-to-the-2019-cpa-exam

CPA Exam Perspective:

https://www.gleim.com/cpa-review/blog/data-analytics-not-the-big-deal-its-made-out-to-be/

https://www.surgentcpareview.com/cpa-blog/how-data-analytics-affect-the-cpa-exam/

Changes on the CPA Exam:

https://www.journalofaccountancy.com/podcast/cpa-exam-changes.html

AICPA: Video:

ttps://players.brightcove.net/1485859309/default\_default/index.html?videoId=5988437676001

*Start at 8:30*

This is a broad discussion:

https://www.aicpa.org/content/dam/aicpa/interestareas/frc/assuranceadvisoryservices/downloadabledocuments/describing-a-set-of-data-and-evaluating-its-integrity.pdf

Future of CPA Data Analytics is now, paper published online in 2019:

https://zicklin.baruch.cuny.edu/wp-content/uploads/sites/10/2019/12/Deniz-Applebaum-Data-Analytics-Knowledge-Required-of-a-CPA.pdf

Broad overview:

<https://harvardmagazine.com/2014/03/why-big-data-is-a-big-deal>

**New Accounting Skills Required for the AI Age**

AI is seen as a technology that will replace a lot of jobs at any skill level. In accounting, AICPA CEO Barry Melancon made an [alarming prediction](https://www.accountingtoday.com/opinion/trends-likely-to-accelerate-merger-mania-among-firms) that the accounting industry could be negatively affected by changes in technology, losing more than 1 million jobs. True or not, accountants will have to adopt this technology, just as they had to adopt the computer or internet.

Tax professionals will have to learn how to implement AI technologies for their daily work activities, for instance, developing the ability to automate their workflow and interpret new kinds of data. They will basically be required to leverage the power of machine learning.

https://www.accountingtoday.com/opinion/how-ai-will-change-the-way-accountants-work?utm\_source=newsletter&utm\_medium=email&utm\_campaign=ACT\_Best\_of\_the\_Week\_LIVE%2B%27-%27%2B07202019&bt\_ee=%2B2bDdyIKkB7stTzOorMriF8RqRcokqcfWW1Ub%2BeIB2JV5R6cIJg3sSZdSedu4P%2Fe&bt\_ts=1563628039553

AI is already present in accounting, and it will grow significantly, as the industry continues to automate and reduce reliance on manual data entry. You can already find AI technology in a number of software applications with benefits including (but not limited to):

* **Scaling up quantity and quality of data analysis:** AI can process huge amounts of data (structured and unstructured), and boosts the scale, scope and rigor of the analysis. Auditors often perform test procedures — AI can literally analyze all the available transactions.
* **Enhancing powers of observation and detection**: AI can extract insights, pick up faint signals and detect more complex patterns in data than humans can.
* **Augmenting cognitive capacity**: Using feedback loops, AI can automatically and instantly learn from errors or new cases and become increasingly smarter over time. It never forgets and continually builds on and deepens the corporate memory.
* **Improving consistency**: AI can be a far more consistent decision-maker. Robots do not get bored, tired, frustrated, moody, lazy, emotional, hungry, thirsty or sick. Machines are not impacted by cycles or fluctuations in biological or physiological states like people. They don’t take vacations or leave of absence either.
* **Mitigating repetitive tasks**: Rather than wasting time with tedious tasks such as data entry and manual review procedures, accountants can focus their efforts on all the work that requires a human touch.
* **Reducing errors**: In a traditional bookkeeping setting, accounting mistakes may go unnoticed. AI can detect errors immediately and ensure your books are always accurate.
* **Clearing invoices faster:** Dealing with payments from multiple invoices can be challenging. Machine learning allows AI to analyze the data and clear out invoices or generate new ones.
* **Accelerating data analysis**: AI can also perform large-scale tasks that would be virtually impossible for humans to complete in a timely manner. Example: AI can analyze the data from every accounting project ever completed in your practice. The technology can then provide valuable insights on how to proceed with a potential project idea most effectively or even recommend that the company scraps it all together. While a human accountant might arrive at the same conclusion, it would take countless hours to go through the data.
* **Real-time audits to ensure compliance**: AI can instantly detect inaccuracies and flag improper submissions in expense reports and travel claims. While carefully reviewing these details would be quite tedious for human accountants, AI tools can learn the company’s policies and analyze data in bulk to ensure that there are no discrepancies.

Robots:

* Panera Bread loses close to 100% of its workers every year.
* For fast-food chains employee turnover runs as high as 130% to 150%, according to industry measures.
* McDonald’s is spending nearly $1 billion in 2019 to add ordering kiosks and other tech to stores.
* Some experts believe it is inevitable that fast food will be the first job sector ruled by robots.

<https://www.cnbc.com/2019/08/29/fast-food-restaurants-in-america-are-losing-100percent-of-workers-every-year.html>

Data Analytics used in the reporting the news:

As reporters and editors find themselves the victims of layoffs at digital publishers and traditional newspaper chains alike, journalism generated by machine is on the rise.

Roughly a third of the content published by Bloomberg News uses some form of automated technology. The system used by the company, Cyborg, is able to assist reporters in churning out thousands of articles on company earnings reports each quarter.

The program can dissect a financial report the moment it appears and spit out an immediate news story that includes the most pertinent facts and figures. And unlike business reporters, who find working on that kind of thing a snooze, it does so without complaint.

Untiring and accurate, Cyborg helps Bloomberg in its race against Reuters, its main rival in the field of quick-twitch business financial journalism, as well as giving it a fighting chance against a more recent player in the information race, hedge funds, which use artificial intelligence to serve their clients fresh facts.

https://www.nytimes.com/2019/02/05/business/media/artificial-intelligence-journalism-robots.html

*A Comprehensive Survey of Data Mining-based Fraud Detection Research:*

Data mining is about finding insights which are statistically reliable, unknown previously, and actionable from data (Elkan, 2001). This data must be available, relevant, adequate, and clean.

Also, the data mining problem must be well-defined, cannot be solved by query and reporting tools, and guided by a data mining process model (Lavrac *et al*, 2004). This survey paper categorises, compares, and summarises from almost all published technical and review articles in automated fraud detection within the last 10 years. It defines the professional fraudster, formalises the main types and subtypes of known fraud, and presents the nature of data evidence collected within affected industries. Within the business context of mining the data to

achieve higher cost savings, this research presents methods and techniques together with their problems. Compared to all related reviews on fraud detection, this survey covers much more

technical articles and is the only one, to the best of our knowledge, which proposes alternative data and solutions from related domains.

[*http://arxiv.org/ftp/arxiv/papers/1009/1009.6119.pdf*](http://arxiv.org/ftp/arxiv/papers/1009/1009.6119.pdf)

*Fraudulent Network Usage*:

Patent for system and method for detecting fraudulent network usage patterns using real-time network monitoring:

<http://www.google.com/patents?hl=en&lr=&vid=USPAT5627886&id=ws0fAAAAEBAJ&oi=fnd&dq=data+driven+fraud+detection&printsec=abstract#v=onepage&q=data%20driven%20fraud%20detection&f=false>

*Detection of Credit Card Fraud*:

The prevention of credit card fraud is an important application for prediction techniques. One major obstacle for using neural network training techniques is the high necessary diagnostic quality: since only one financial transaction in a thousand is invalid no prediction success less than 99.9% is acceptable. Because of these credit card transaction requirements, completely new concepts had to be developed and tested on real credit card data. This paper shows how advanced data mining techniques and a neural network algorithm can be combined successfully to obtain a high fraud coverage combined with a low false alarm rate.

<http://pdf.aminer.org/000/244/982/credit_card_fraud_detection_with_a_neural_network.pdf>

*Data Analysis*:

The outlier detection problem has important applications in the field of fraud detection, network robustness analysis, and intrusion detection. Most such applications are high dimensional domains in which the data can contain hundreds of dimensions. Many recent algorithms use concepts of proximity in order to find outliers based on their relationship to the rest of the data. However, in high dimensional space, the data is sparse and the notion of proximity fails to retain its meaningfulness. In fact, the sparsity of high dimensional data implies that every point is an almost equally good outlier from the perspective of proximity-based definitions. Consequently, for high dimensional data, the notion of finding meaningful outliers becomes substantially more complex and non-obvious. In this paper, we discuss new techniques for outlier detection which find the outliers by studying the behavior of projections from the data set.

<http://ftp.cse.buffalo.edu/users/azhang/disc/SIGMOD/pdf-files/037/172-outlier.pdf>

*System Fraud Analysis*:

Fraudsters aim to use services without paying or illicitly benefit from the service in other ways, causing service providers financial damage. We report an experiment aimed at generating synthetic test data for fraud detection in an IP based video-on-demand service. The data generation verifies a methodology previously developed by the present authors [E. Lundin et al., (2002)] that ensures that important statistical properties of the authentic data are preserved by using authentic normal data and fraud as a seed for generating synthetic data. This enables us to create realistic behavior profiles for users and attackers. The data is used to train the fraud detection system itself, thus creating the necessary adaptation of the system to a specific environment. Here we aim to verify the usability and applicability of the synthetic data, by using them to train a fraud detection system. The system is then exposed to a set of authentic data to measure parameters such as detection capability and false alarm rate as well as to a corresponding set of synthetic data, and the results are compared.

<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.4.4860&rep=rep1&type=pdf>

*Financial Statement Fraud Analysis:*

Data mining techniques have been used enormously by the researchers’ community in detecting financial statement fraud. Most of the research in this direction has used the numbers (quantitative information) i.e. financial ratios present in the financial statements for detecting fraud. There is very little or no research on the analysis of text such as auditor’s comments or notes present in published reports. In this study we propose a text mining approach for detecting financial statement fraud by analyzing the hidden clues in the qualitative information (text) present in financial statements.

http://thesai.org/Downloads/Volume3No12/Paper\_30-Financial\_Statement\_Fraud\_Detection\_using\_Text\_Mining.pdf

**How to replace an employee**

RPA takes the robot out of the human. The average knowledge worker employed on a back office process has a lot of repetitive, routine tasks that are dreary and uninteresting. RPA is a type of software that mimics the activity of a human being in carrying out a task within a process. It can do repetitive stuff more quickly, accurately, and tirelessly than humans, freeing them to do other tasks requiring human strengths such as emotional intelligence, reasoning, judgment, and interaction with the customer

There are four streams of RPA. The first is a highly customized software that will work only with certain types of process in, say, accounting and finance. The more general streams I describe in terms of a three-lane motorway. The slow lane is what we call screen scraping or web scraping. A user might be collecting data, synthesizing it, and putting it into some sort of document on a desktop. You automate as much of that as possible. The second lane in terms of power is a self-development kit where a template is provided and specialist programmers design the robot. That’s usually customized for a specific organization. The fast lane is enterprise/enterprise-safe software that can be scaled and is reusable.

**https://www.mckinsey.com/industries/financial-services/our-insights/the-value-of-robotic-process-automation**

**What is the IQ worker replacement?**

German psychologist William Stern. As the article states, the foundation of IQ testing is to gauge human pattern-matching aptitudes in the following three areas:

* Logical: identifying patterns in sequences of concepts
* Mathematical: identifying patterns in sequences of numbers
* Linguistic: identifying patterns in words, primarily focused on semantic patterns such as analogies, classifications, synonyms and antonyms

These cognitive tasks may not be the only types of intelligence worth considering.

As we consider advances in machine learning algorithms, we recognize that it’s only a matter of time before computers will ace practically any standard IQ test we could throw at them. With their multilayered ability to find patterns in linguistic, mathematical and other patterns within complex content corpora, these and other cognitive-computing algorithms will indeed grow smarter—and smarter, and smarter still.

As a case in point, take reports on advances by researchers in China, who have developed a [deep-learning implementation](http://www.technologyreview.com/view/538431/deep-learning-machine-beats-humans-in-iq-test/) that outperforms average humans in answering verbal reasoning questions. They measured their algorithm’s performance against that of a crowdsourced set of humans participating via Amazon Mechanical Turk.

<https://www.ibmbigdatahub.com/blog/measuring-artificial-intelligence-quotient>

This is how the jobs are being replaced:

It is this last category that has interested Huazheng Wang and pals at the University of Science and Technology of China and Bin Gao and buddies at Microsoft Research in Beijing. Computers have never been good at these. Pose a verbal reasoning question to a natural language processing machine and its performance will be poor, much worse than the average human ability.

Today, that changes thanks to Huazheng and pals who have built a deep learning machine that outperforms the average human ability to answer verbal reasoning questions for the first time.

In recent, years, computer scientists have used data mining techniques to analyze huge corpuses of texts to find the links between words they contain. In particular, this gives them a handle on the statistics of word patterns, such as how often a particular word appears near other words. From this it is possible to work out how words relate to each other, albeit in a huge parameter space.

The end result is that words can be thought of as vectors in this high-dimensional parameter space. the advantage is that they can then be treated mathematically: compared, added, subtracted like other vectors. This leads to vector relations like this one: king – man + woman = queen.

This approach has been hugely successful.  Google uses it for automatic language translation by assuming that word sequences in different language represented by similar vectors are equivalent in meaning. So they are translations of each other.

But this approach has a well-known shortcoming: it assumes that each word has a single meaning represented by a single vector. Not only is that often not the case, verbal tests tend to focus on words with more than one meaning as a way of making questions harder.

Huazheng and pals tackle this by taking each word and looking for other words that often appear nearby in a large corpus of text. They then use an algorithm to see how these words are clustered. The final step is to look up the different meanings of a word in a dictionary and then to match the clusters to each meaning.

This can be done automatically because the dictionary definition includes sample sentences in which the word is used in each different way. So by calculating the vector representation of these sentences and comparing them to the vector representation in each cluster, it is possible to match them.

The overall result is a way of recognizing the multiple different senses that some words can have.

Huazheng and pals have another trick up their sleeve to make it easier for a computer to answer verbal reasoning questions. This comes about because these questions fall into several categories that require slightly different approaches to solve.

So their idea is to start by identifying the category of each question so that the computer then knows which answering strategy it should employ. This is straightforward since the questions in each category have similar structures.

So questions that involve analogies are like these:

Isotherm is to temperature as isobar is to? (i) atmosphere, (ii) wind, (iii) pressure, (iv) latitude, (v) current.

and

Identify two words (one from each set of brackets) that form a connection (analogy) when paired with the words in capitals: CHAPTER (book, verse, read), ACT (stage, audience, play).

Word classification questions are like this:

Which is the odd one out? (i) calm, (ii) quiet, (iii) relaxed, (iv) serene, (v) unruffled.

And questions looking for synonyms and antonyms are like these:

Which word is closest to IRRATIONAL? (i) intransigent, (ii) irredeemable, (iii) unsafe, (iv) lost, (v) nonsensical.

And

Which word is most opposite to MUSICAL? (i) discordant, (ii) loud, (iii) lyrical, (iv) verbal, (v) euphonious.

Spotting each type of question is relatively straightforward for a machine learning algorithm, given enough to examples to learn from. And this is exactly how Huazheng and co do it.

Having identified the type of question, Huazheng and buddies then devise an algorithm for solving each one using the standard vector methods but also the multi-sense upgrade they’ve developed.

They compare this deep learning technique with other algorithmic approaches to verbal reasoning tests and also with the ability of humans to do it. For this, they posed the questions to 200 humans gathered via Amazon’s Mechanical Turk crowdsourcing facility along with basic information about their ages and educational background.

And the results are impressive. “To our surprise, the average performance of human beings is a little lower than that of our proposed method,” they say.

Human performance on these tests tends to correlate with educational background. So people with a high school education tend to do least well, while those with a bachelor’s degree do better and those with a doctorate perform best. “Our model can reach the intelligence level between the people with the bachelor degrees and those with the master degrees,” say Huazheng and co.

That’s fascinating work that reveals the potential of deep learning techniques. Huazheng and co are clearly optimistic about future developments. “With appropriate uses of the deep learning technologies, we could be a further step closer to the true human intelligence.”

Deep learning techniques are currently sweeping through computer science like wildfire and the revolution they are creating is still in its early stages. There’s no telling where this revolution will take us but one thing is for sure: William Stern would be amazed.

<https://www.technologyreview.com/s/538431/deep-learning-machine-beats-humans-in-iq-test/>

**Break each job down into a specific component and then write a program to perform that component.**

This article show how employers will break apart certain duties and then make then automate the duties.

<https://www.mckinsey.com/featured-insights/digital-disruption/whats-now-and-next-in-analytics-ai-and-automation#section%202>

### AI Improves Significantly

“The results so far indicate that the artificial intelligence systems produced by Google, Baidu, and others have significantly improved over the past two years but still have certain gaps as compared with even a six-year-old child,” [according to the research](https://arxiv.org/ftp/arxiv/papers/1709/1709.10242.pdf).

While AI’s IQ scores are still dwarfed by that of the average human adult, Google and Microsoft have all seen their [research and development (R&D)](https://www.investopedia.com/terms/r/randd.asp) pay off in increasingly intelligent [bots](https://www.investopedia.com/terms/c/chatbot.asp). In 2014, Google’s IQ score was just 26.5, while Microsoft’s was 13.5.

https://www.cfo.com/it-value/2019/04/advanced-technologies-are-taking-over-finance/

Digital Analytics & A.I

Alphabet Inc.’s ([GOOGL](https://www.investopedia.com/markets/quote?tvwidgetsymbol=googl)) Google has developed [artificial intelligence](https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp) that has much more brain power than Apple Inc.’s ([AAPL](https://www.investopedia.com/markets/quote?tvwidgetsymbol=aapl)) virtual assistant Siri, according to a new academic paper. Based on data compiled by a trio of Chinese researchers who have developed an IQ test for AI, Google’s score significantly outperforms Siri and Bing—yet with an IQ less than that of a 6-year-old. (See also: [Artificial Intelligence Will Add $15.7 Trillion to the Global Economy: PwC.](https://www.investopedia.com/terms/a/artificial-intelligence-ai.asp))

Researchers Feng Liu, Yong Shi and Yin Liu carried out tests throughout 2016, which ranked Google’s AI IQ at 47.28, just shy of the average IQ they found for a human 6-year-old: 55.5. Siri’s IQ fell well below at 23.9, which was also lower than Microsoft Corp.’s ([MSFT](https://www.investopedia.com/markets/quote?tvwidgetsymbol=msft)) Bing and Baidu, at 31.98 and 32.92, respectively. For comparison’s sake, the average 18-year-old has an IQ of 97, according to the study.

**Augmented Reality**

Interest in augmented reality (AR) is exploding as innovators explore the business relevance and roles AR can play in workforce enablement and customer experience and interaction. This emerging technology holds tremendous promise for changing the way businesses operate.

*Video*: https://www.youtube.com/watch?v=vQtwWzfzKXI

Deepfakes: “Deepfakes” refers to media that has been altered by artificial intelligence to make it appear that a person is doing or saying something that, in fact, that person has never done or said.

Video: <https://www.youtube.com/watch?time_continue=3&v=cQ54GDm1eL0&feature=emb_title>

https://www.fastcompany.com/90379001/criminals-are-using-deepfakes-to-impersonate-ceos