

Major AI and ML Breakthroughs in 2018 and Trends to Look out for in 2019

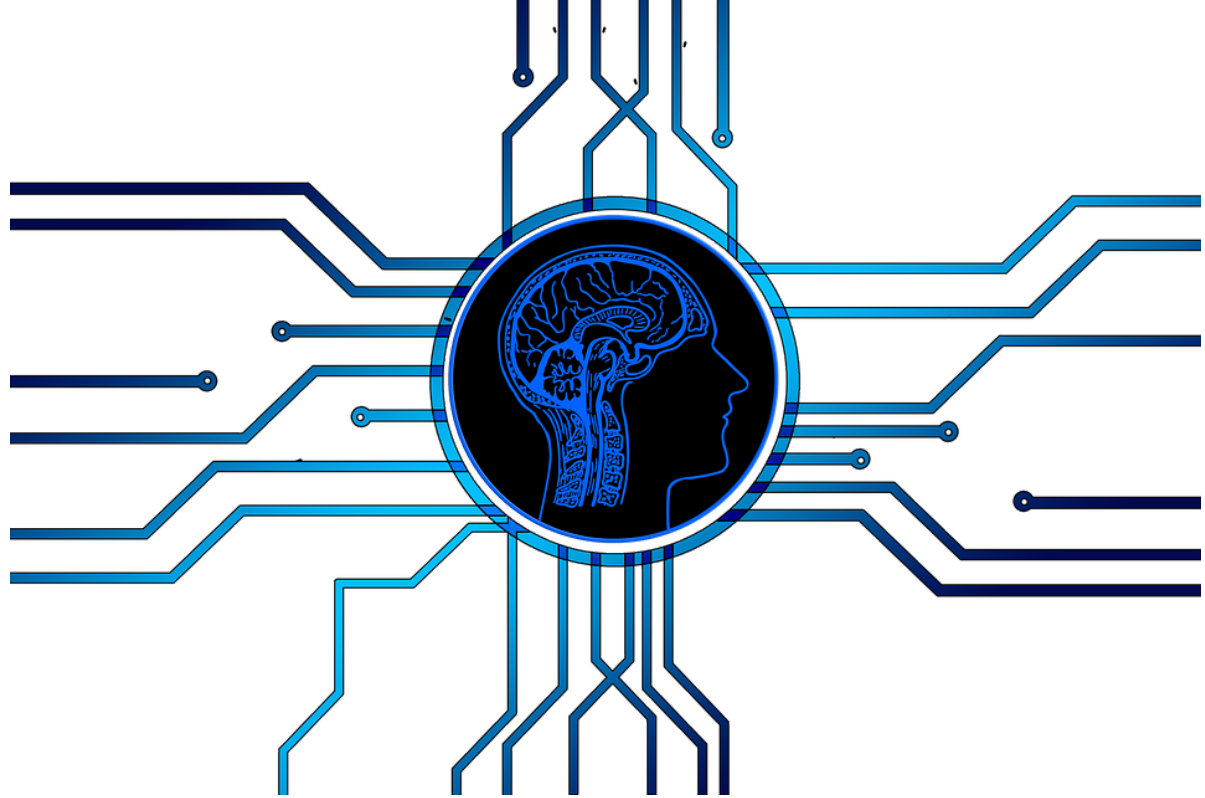
Introduction

The last few years have been a dream run for Artificial Intelligence enthusiasts and machine learning professionals. These technologies have evolved from being a niche to becoming mainstream, and are impacting millions of lives today. Countries now have dedicated AI ministers and budgets to make sure they stay relevant in this race.

The same has been true for a data science professional. A few years back – you would have been comfortable knowing a few tools and techniques. Not anymore! There is so much happening in this domain and so much to keep pace with – it feels mind boggling at times.

This is why I thought of taking a step back and looking at the developments in some of the key areas in Artificial Intelligence from a data science practitioners' perspective. What were these breakthroughs? What happened in 2018 and what can be expected in 2019? Read this article to find out!

P.S. As with any forecasts, these are my takes. These are based on me trying to connect the dots. If you have a different perspective – I would love to hear it. Do let me know what you think might change in 2019.



Areas we'll cover in this article

- Natural Language Processing (NLP)
- Computer Vision
- Tools and Libraries
- Reinforcement Learning
- AI for Good – A Move Towards Ethical AI

Natural Language Processing (NLP)

The best part about ULMFiT and the subsequent frameworks we'll see soon? You don't need to train models from scratch! These researchers have done the hard bit for you – take their learning and apply it in your own projects. ULMFiT outperformed state-of-the-art methods in six text classification tasks.

You can read [this excellent tutorial](#) by Prateek Joshi on how to get started with ULMFiT for any text classification problem.

ELMo

Want to take a guess at what ELMo stands for? It's short for Embeddings from Language Models. Pretty creative, eh? Apart from its name resembling the famous Sesame Street character, ELMo grabbed the attention of the ML community as soon as it was released.

ELMo uses language models to obtain embeddings for each word while also considering the context in which the word fits into the sentence or paragraph. Context is such a crucial aspect of NLP that most people failed to grasp before. ELMo uses bi-directional LSTMs to create the embeddings. Don't worry if that sounds like a mouthful – check out [this article](#) to get a really simple overview of what LSTMs are and how they work.

Like ULMFiT, ELMo significantly improves the performance of a wide variety of NLP tasks, like sentiment analysis and question answering. Read more about it [here](#).

Google's BERT

Quite a few experts have claimed that the release of BERT marks a new era in NLP. Following ULMFiT and ELMo, BERT really blew away the competition with its performance. As the [original paper](#) states, "BERT is conceptually simple and empirically powerful".

BERT obtained state-of-the-art results on 11 (yes, 11!) NLP tasks. Check out their results on the SQuAD benchmark:

SQuAD v1.1 Leaderboard (Oct 8th 2018) Test EM Test F1

1st Place Ensemble – BERT	87.4	93.2
---------------------------	------	------

SQuAD v1.1 Leaderboard (Oct 8th 2018) Test EM Test F1

2nd Place Ensemble – nlnet	86.0	91.7
1st Place Single Model – BERT	85.1	91.8
2nd Place Single Model – nlnet	83.5	90.1

Interested in getting started? You can use either the [PyTorch implementation](#) or Google's own [TensorFlow code](#) to try and replicate the results on your own machine.

I'm fairly certain you are wondering what BERT stands for at this point.

😊 It's Bidirectional Encoder Representations from Transformers. Full marks if you got it right the first time.

Facebook's PyText

How could Facebook stay out of the race? They have open-sourced their own deep learning NLP framework called PyText. It was released earlier this week so I'm still to experiment with it, but the early reviews are extremely promising. According to research published by FB, PyText has led to a 10% increase in accuracy of conversational models and reduced the training time as well.

PyText is actually behind a few of Facebook's own products like the FB Messenger. So working on this adds some real-world value to your own portfolio (apart from the invaluable knowledge you'll gain obviously).

You can try it out yourself by downloading the code from [this GitHub repo](#).

Google Duplex

If you haven't heard of Google Duplex yet, where have you been?! Sundar Pichai knocked it out of the park with this demo and it has been in the headlines ever since:

Google I/O 2018 Google's Duplex Demo Stole the Show



Since this is a Google product, there's a slim chance of them open sourcing the code behind it. But wow! That's a pretty awesome audio processing application to showcase. Of course it raises a lot of ethical and privacy questions, but that's a discussion for later in this article. For now, just revel in how far we have come with ML in recent years.

NLP Trends to Expect in 2019

Who better than Sebastian Ruder himself to provide a handle on where NLP is headed in 2019? Here are his thoughts:

1. **Pretrained language model embeddings will become ubiquitous**; it will be rare to have a state-of-the-art model that is not using them
2. We'll see **pretrained representations that can encode specialized information** which is complementary to language model embeddings. We will be able to combine different types of pretrained representations depending on the requirements of the task
3. We'll see more work on **multilingual applications and cross-lingual models**. In particular, building on cross-lingual word embeddings, we will see the emergence of

Computer Vision



This is easily the most popular field right now in the deep learning space. I feel like we have plucked the low-hanging fruits of computer vision to quite an extent and are already in the refining stage. Whether it's image or video, we have seen a plethora of frameworks and libraries that have made computer vision tasks a breeze.

We at Analytics Vidhya spent a lot of time this year working on democratizing these concepts. Check out our [computer vision specific articles here](#), covering topics from object detection in videos and images to lists of pretrained models to get your deep learning journey started.

Here's my pick of the best developments we saw in CV this year.

And if you're curious about this wonderful field (actually going to become one of the hottest jobs in the industry soon), then go ahead and start your journey with our '[Computer Vision using Deep Learning](#)' course.

The Release of BigGANs

Ian Goodfellow designed GANs in 2014, and the concept has spawned multiple and diverse applications since. Year after year we see the

original concept being tweaked to fit a practical use case. But one thing has remained fairly consistent till this year – images generated by machines were fairly easy to spot. There would always be some inconsistency in the frame which made the distinction fairly obvious.

But that boundary has started to seep away in recent months. And with the creation of [BigGANs](#), that boundary could be removed permanently. Check out the below images generated using this method:



Unless you take a microscope to it, you won't be able to tell if there's anything wrong with that collection. Concerning or exciting? I'll leave that up to you, but there's no doubt GANs are changing the way we perceive digital images (and videos).

For the data scientists out there, these models were trained on the ImageNet dataset first and then the JFT-300M data to showcase that these models transfer well from one set to the other. I would also to direct you to the [GAN Dissection page](#) – a really cool way to visualize and understand GANs.

Fast.ai's Model being Trained on ImageNet in 18 Minutes

This was a really cool development. There is a very common belief that you need a ton of data along with heavy computational resources to perform proper deep learning tasks. That includes training a model from scratch on the ImageNet dataset. I understand that perception – most of us thought the same before a few folks at fast.ai found a way to prove all of us wrong.

Their model gave an accuracy of 93% in an impressive 18 minutes timeframe. The hardware they used, detailed in their [blog post](#), contained 16 public AWS cloud instances, each with 8 NVIDIA V100 GPUs. They built the algorithm using the fastai and PyTorch libraries.

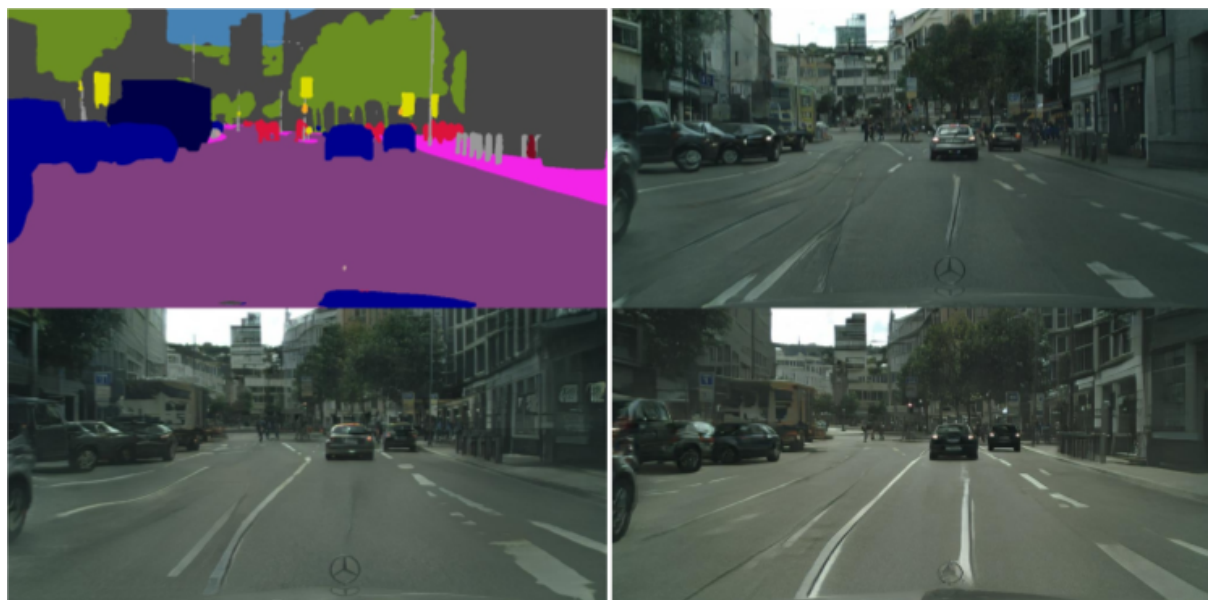


The total cost of putting the whole thing together came out to be just \$40! **Jeremy has described their approach, including techniques, in much more detail [here](#).** A win for everyone!

NVIDIA's vid2vid technique

Image processing has come leaps and bounds in the last 4-5 years, but what about video? Translating methods from a static frame to a dynamic one has proved to be a little tougher than most imagined. Can you take a video sequence and predict what will happen in the next frame? It had been explored before but the published research had been vague, at best.

NVIDIA decided to open source their approach earlier this year, and it was met with widespread praise. The goal of their vid2vid approach is to learn a mapping function from a given input video in order to produce an output video which depicts the contents of the input video with incredible precision.



You can try out their PyTorch implementation available on their [GitHub here](#).

Computer Vision Trends to Expect in 2019

Like I mentioned earlier, we might see modifications rather than inventions in 2019. It might feel like more of the same – self-driving cars, facial recognition algorithms, virtual reality, etc. Feel free to disagree with me here and add your point of view – I would love to know what else we can expect next year that we haven't already seen.

Drones, pending political and government approvals, might finally get the green light in the United States (India is far behind there).

Personally, I would like to see a lot of the research being implemented in real-world scenarios. Conferences like CVPR and ICML portray the latest in this field but how close are those projects to being used in reality?

Visual question answering and visual dialog systems could finally make their long-awaited debut soon. These systems lack the ability to generalize but the expectation is that we'll see an integrated multi-modal approach soon.

Self-supervised learning came to the forefront this year. I can bet on that being used in far more studies next year. It's a really cool line of learning – the labels are directly determined from the data we input, rather than wasting time labelling images manually. Fingers crossed!

Tools and Libraries

This section will appeal to all data science professionals. Tools and libraries are the bread and butter of data scientists. I have been part a part of plenty of debates about which tool is the best, which framework supersedes the other, which library is the epitome of economical computations, etc. I'm sure quite a lot of you will be able to relate to this as well.

But one thing we can all agree on – we need to be on top of the latest tools in the field, or risk being left behind. The pace with which Python has overtaken everything else and planted itself as the industry leader

is example enough of this. Of course a lot of this comes down to subjective choices (what tool is your organization using, how feasible is it to switch from the current framework to a new one, etc.), but if you aren't even considering the state-of-the-art out there, then I implore you to start NOW.

So what made the headlines this year? Let's find out!

PyTorch 1.0

What's all the hype about PyTorch? I've mentioned it multiple times already in this article (and you'll see more instances later). I'll leave it to my colleague Faizan Shaikh to [acquaint you with the framework](#).



That's one of my favorite deep learning articles on AV – a must-read! Given how slow TensorFlow can be at times, it opened the door for PyTorch to capture the deep learning market in double-quick time. Most of the code that I see open sourced on GitHub is a PyTorch implementation of the concept. It's not a coincidence – PyTorch is super flexible and the latest version (v1.0) already powers many Facebook products and services at scale, including performing 6 billion text translations a day.

PyTorch's adoption rate is only going to go up in 2019 so now is as good a time as any to get on board.

AutoML – Automated Machine Learning

Automated machine learning (or AutoML) has been gradually making inroads in the last couple of years. Companies like RapidMiner, KNIME, DataRobot and H2O.ai have released excellent products showcasing the immense potential of this service.

Can you imagine working on a ML project where you only need to work with a drag-and-drop interface without coding? It's a scenario that's not too far off in the future. But apart from these companies, there was a significant release in the ML/DL space – **Auto Keras!**



It's an open source library for performing AutoML tasks. The idea behind it is to make deep learning accessible to domain experts who perhaps don't have a ML background. **Make sure you check it out [here](#).** It is primed to make a huge run in the coming years.

TensorFlow.js – Deep Learning in the Browser

We've been building and designing machine learning and deep learning models in our favorite IDEs and notebooks since we got into this line of work. How about taking a step out and trying something different? Yes, I'm talking about performing deep learning in your web browser itself!



This is now a reality thanks to the release of [TensorFlow.js](#). That link has a few demos as well which demonstrate how cool this open source concept is. There are primarily three advantages/features of TensorFlow.js:

- Develop and deploy machine learning models with JavaScript

- Run pre-existing TensorFlow models in your browser
- Retrain pre-existing models

AutoML Trends to Expect in 2019

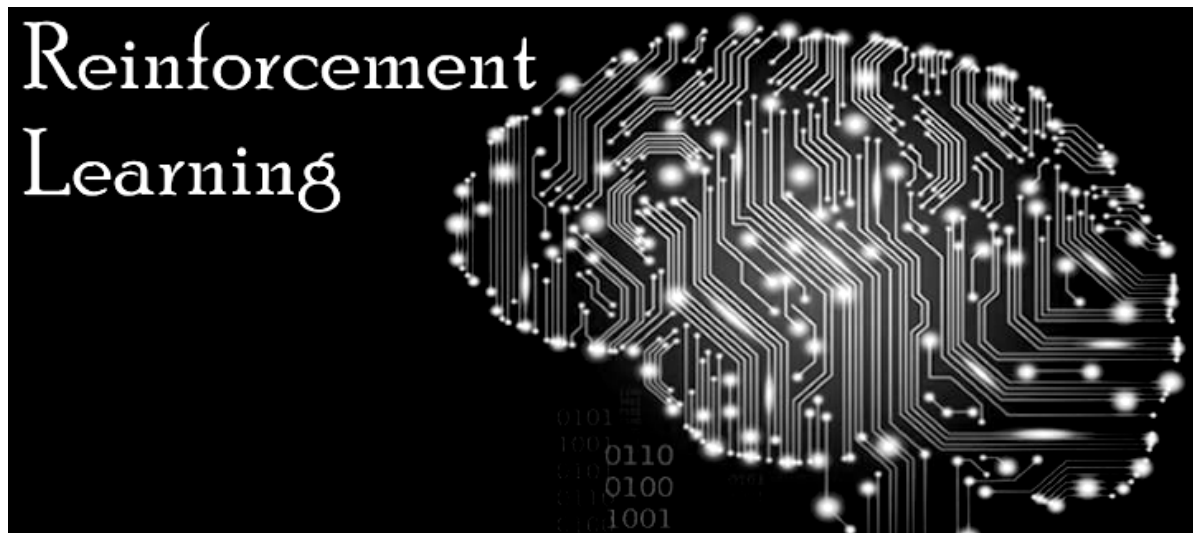
I wanted to focus particularly on AutoML in this thread. Why? Because I feel it's going to be a real-game changer in the data science space in the next few years. But don't just take my word for it! Here's H2O.ai's [Marios Michailidis](#), Kaggle Grandmaster, with his view of what to expect from AutoML in 2019:

Machine learning continues its march into being one of the most important trends of the future – of where the world is going towards to. This expansion has increased the demand for skilled applications in this space. Given its growth, it is imperative that automation is the key into utilising the data science resources as best as possible. The applications are limitless: Credit, insurance, fraud, computer vision, acoustics, sensors, recommenders, forecasting, NLP – you name it. It is a privilege to be working in this space. The trends that will continue being important can be defined as:

1. Providing smart visualisations and insights to help describe and understand the data
2. Finding/building/extracting better features for a given dataset
3. Building more powerful/smarter predictive models – quickly
4. Bridging the gap between black box modelling and productionisation of these models with machine learning interpretability (mli)

5. Facilitating the productionisation of these models

Reinforcement Learning



If I had to pick one field where I want to see more penetration, it would be reinforcement learning. Apart from the occasional headlines we see at irregular intervals, there hasn't yet been a game-changing breakthrough. The general perception I have seen in the community is that it's too math-heavy and there are no real industry applications to work on.

While this is true to a certain extent, I would love to see more practical use cases coming out of RL next year. In my monthly GitHub and Reddit series, I tend to keep at least one repository or discussion on RL to at least foster a discussion around the topic. This might well be the next big thing to come out of all that research.

OpenAI have released a really helpful toolkit to get beginners started with the field, which I have mentioned below. You can also check out [this beginner-friendly introduction](#) on the topic (it has been super helpful for me).

If there's anything I have missed, would love to hear your thoughts on it.

OpenAI's Spinning Up in Deep Reinforcement Learning



If research in RL has been slow, the educational material around it has been minimal (at best). But true to their word, OpenAI have open sourced some awesome material on the subject. They are calling this project ‘Spinning Up in Deep RL’ and you can read all about it [here](#).

It’s actually quite a comprehensive list of resources on RL and they have attempted to keep the code and explanations as simple as possible. There is quite a lot of material which includes things like RL terminologies, how to grow into an RL research role, a list of important papers, a supremely well-documented code repository, and even a few exercises to get you started.

No more procrastinating now – if you were planning to get started with RL, your time has come!

Dopamine by Google

To accelerate research and get the community more involved in reinforcement learning, the Google AI team has open sourced Dopamine, a TensorFlow framework that aims to create research by making it more flexible and reproducible.



You can find the entire training data along with the TensorFlow code (just 15 Python notebooks!) on [this GitHub repository](#). Here’s the

perfect platform for performing easy experiments in a controlled and flexible environment. Sounds like a dream for any data scientist.

Reinforcement Learning Trends to Expect in 2019

Xander Steenbrugge, speaker at DataHack Summit 2018 and founder of the ArxivInsights channel, is quite the expert in reinforcement learning. Here are his thoughts on the current state of RL and what to expect in 2019:

- I currently see three major problems in the domain of RL:
 1. **Sample complexity** (the amount of experience an agent needs to see/gather in order to learn)
 2. **Generalization** and transfer learning (Train on task A, test on related task B)
 3. **Hierarchical RL** (automatic subgoal decomposition)
- I believe that the first two problems can be addressed with a similar set of techniques all related to **unsupervised representation learning**. Currently in RL, we are training Deep Neural nets that map from raw input space (eg Pixels) to actions in an end-to-end manner (eg. with Backpropagation) using sparse reward signals (eg the score of an Atari game or the success of a robotic grasp). The problem here is that:
 - It takes a really long time to actually “grow” useful feature detectors because the signal-to-noise ratio is very low. RL basically starts with random actions until it is lucky enough to stumble upon a reward and then needs to figure out how that specific reward was actually

caused. Further exploration is either hardcoded (epsilon-greedy exploration) or encouraged with techniques like [curiosity-driven-exploration](#). This is not efficient and this leads to problem 1.

- Secondly, these deep NN architectures are known to be very prone to overfitting, and in RL we generally tend to test agents on the training data –> overfitting is actually encouraged in this paradigm.
- A possible path forward that I am very enthusiastic about is to leverage unsupervised representation learning ([autoencoders](#), [VAE's](#), GANs, ...) to transform a messy, high-dimensional input space (eg Pixels) into a lower-dimensional 'conceptual' space that has certain desirable properties such as:
 - Linearity, disentanglement, robustness to noise, ...
- Once you can map Pixels into such a useful latent space, learning suddenly becomes much easier/faster (problem 1.) and you also hope that policies learned in this space will have much stronger [generalization](#) because of the properties mentioned above (problem 2.)
- I am not an expert on the Hierarchy problem, but everything mentioned above also applies here: it's easier to solve a complicated hierarchical task in latent space than it is in raw input space.

BONUS: Check out [Xander's video](#) about overcoming sparse rewards in Deep RL (the first challenge highlighted above).

1. Sample complexity will continue to improve due to adding more and more auxiliary learning tasks that augment the sparse, extrinsic reward signal (things like curiosity driven exploration, autoencoder-style pretraining, disentangling causal factors in the environment, ...). This will work especially well in very sparse reward environments (such as the recent [Go-explore results](#) on Montezuma's revenge)
2. Because of this, training systems directly in the physical world will become more and more feasible (instead of current applications that are mostly trained in simulated environments and then use [domain randomization](#) to transfer to the real world.) I predict that 2019 will bring the first truly impressive robotics demo's that are only possible using Deep Learning approaches and cannot be hardcoded / human engineered (unlike most demo's we have seen so far)
3. Following the major success of Deep RL in the AlphaGo story (especially with the recent [AlphaFold](#) results), I believe RL will gradually start delivering actual business applications that create real-world value outside of the academic space. This will initially be limited to applications where accurate simulators are available to do large-scale, virtual training of these agents (eg drug discovery, electronic-chip architecture optimization, vehicle & package routing, ...)
4. As has already started to happen (see [here](#) or [here](#)) there will be a general shift in RL

development where testing an agent on the training data will no longer be considered 'allowed'. Generalization metrics will become core, just as is the case for supervised learning methods

AI for Good – A Move Towards Ethical AI

Imagine a world ruled by algorithms that dictate every action humans take. Not exactly a rosy scenario, is it? Ethics in AI is a topic we at Analytics Vidhya have always been keen to talk about. It becomes bogged down amid all the technical discussions when it should be considered along with those topics.

Quite a few organizations were left with egg on their face this year with Facebook's Cambridge Analytica scandal and Google's internal rife about designing weapons headlining the list of scandals. But all of this led to the big tech companies penning down charters and guidelines they intend to follow.

There isn't one out-of-the-box solution or one size fits all solution to handling the ethical aspect of AI. It requires a nuanced approach combined with a structured path put forward by the leadership. Let's see a couple of major moves that shook the landscape earlier this year.

Campaigns by Google and Microsoft

It was heartening to see the big corporations putting emphasis on this side of AI (even though the road that led to this point wasn't pretty). I want to direct your attention to the guidelines and principles released by a couple of these companies:

- [Google's AI Principles](#)
- [Microsoft's AI Principles](#)

These all essentially talk about fairness in AI and when and where to draw the line. Always a good idea to reference them when you're starting a new AI based project.

How GDPR has Changed the Game

GDPR, or the General Data Protection Regulation, has definitely had an impact on the way data is collected for building AI applications. GDPR came into play to ensure users have more control over their data (what information is collected and shared about them).

So how does that affect AI? Well, if the data scientist does not have data (or enough of it), building any model becomes a non-starter. This has certainly put a spanner in the works of how social platforms and other sites used to work. GDPR will make for a fascinating case study down the line but for now, it has limited the usefulness of AI for a lot of platforms.

Ethical AI Trends to Expect in 2019

This is a bit of a grey field. Like I mentioned, there's no one solution to it. We have to come together as a community to integrate ethics within AI projects. How can we make that happen? As Analytics Vidhya's Founder and CEO Kunal Jain highlighted in his talk at DataHack Summit 2018, we will need to pen down a framework which others can follow.

I expect to see new roles being added in organizations that primarily deal with ethical AI. Corporate best practices will need to be re-structured and governance approaches re-drawn as AI becomes central to the company's vision. I also expect the Government to play a more active role in this regard with new or modified policies coming into play. 2019 will be a very interesting year, indeed.

End Notes

Impactful – the only word that succinctly describes the amazing developments in 2018. I've become an avid user of ULMFiT this year and I'm looking forward to exploring BERT soon. Exciting times, indeed.

I would love to hear from you as well! What developments did you find the most useful? Are you working on any project using the frameworks/tools/concepts we saw in this article? And what are your predictions for the coming year? I look forward to hearing your thoughts and ideas in the comments section below.

You can also read this article on Analytics Vidhya's Android APP



Like this:

Like Loading...

Viewed using [Just Read](#)