POSTSECONDARY SOLUTIONS:
AI AND MACHINE LEARNING

WITH POLYOLYGON

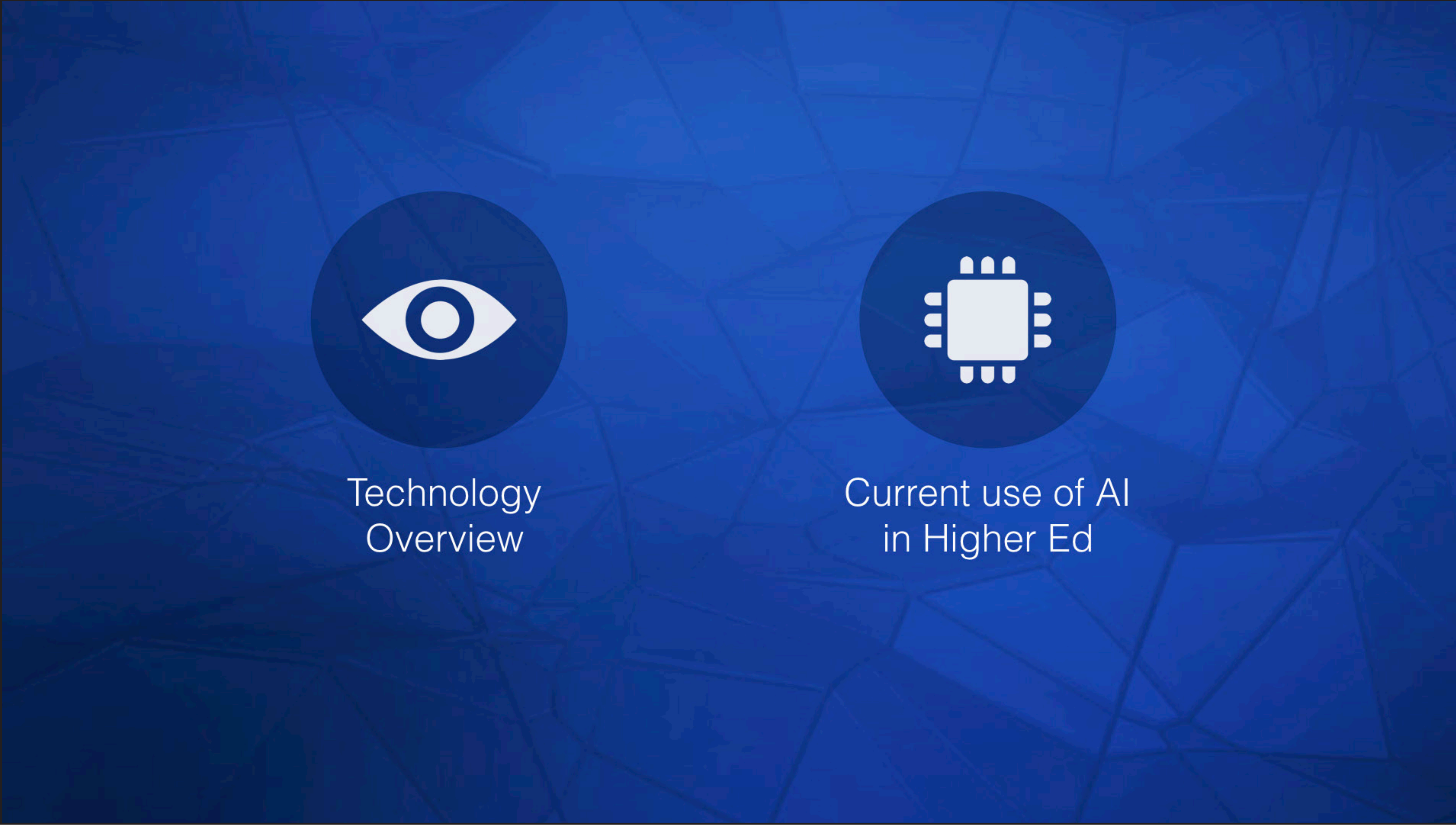
BILL & MELINDA GATES foundation

- This is the second of several nascent technology presentations we'll be doing for Post Secondary.
- The first topic, AR & VR was more "Blue Skies" in term of current Higher Ed utility but we'll find today's topic on AI & Machine Learning quite the opposite.

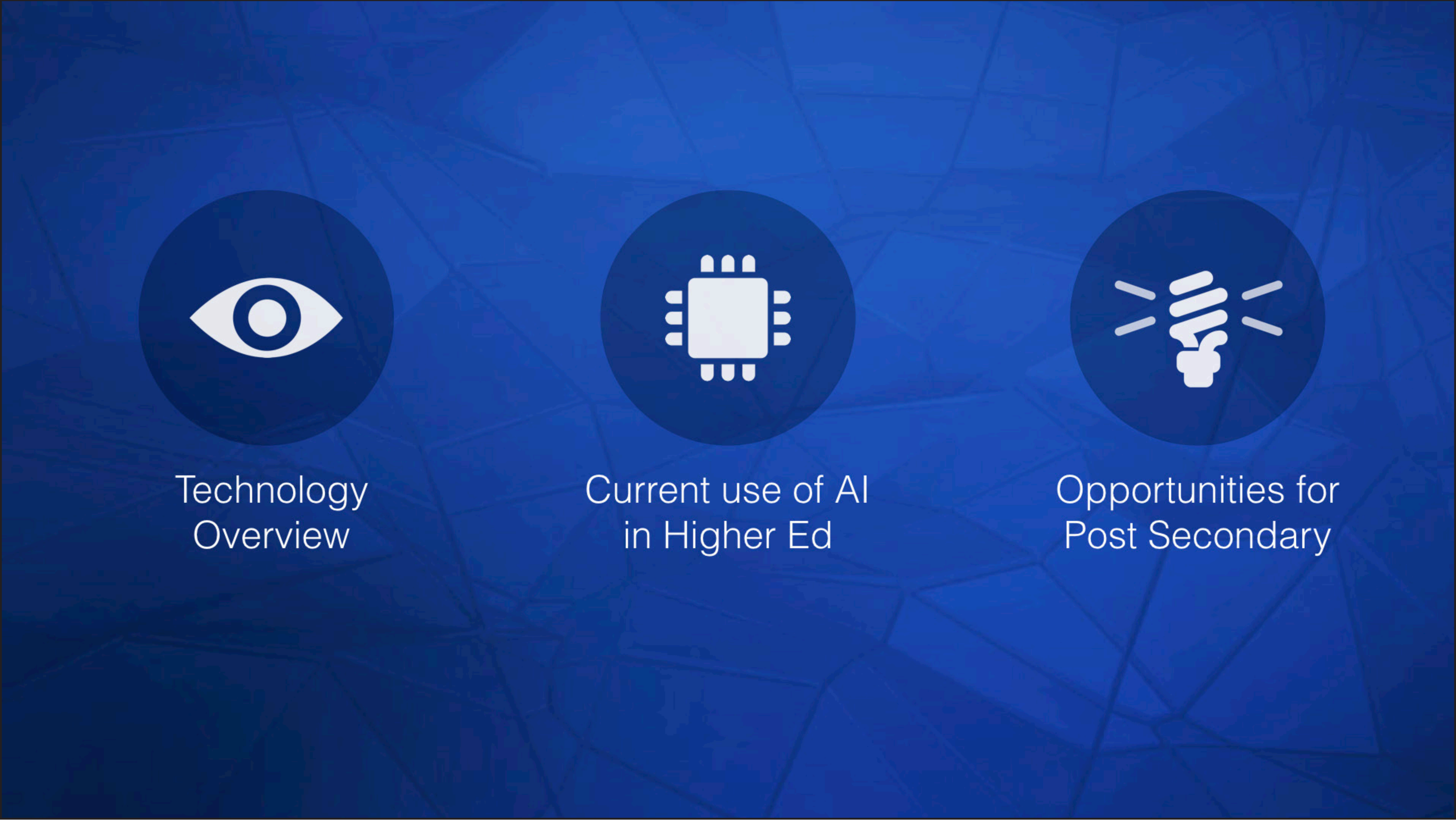
OBJECTIVE

- Before we dive in let’s restate what we’re aiming to take away from these technology presentations…

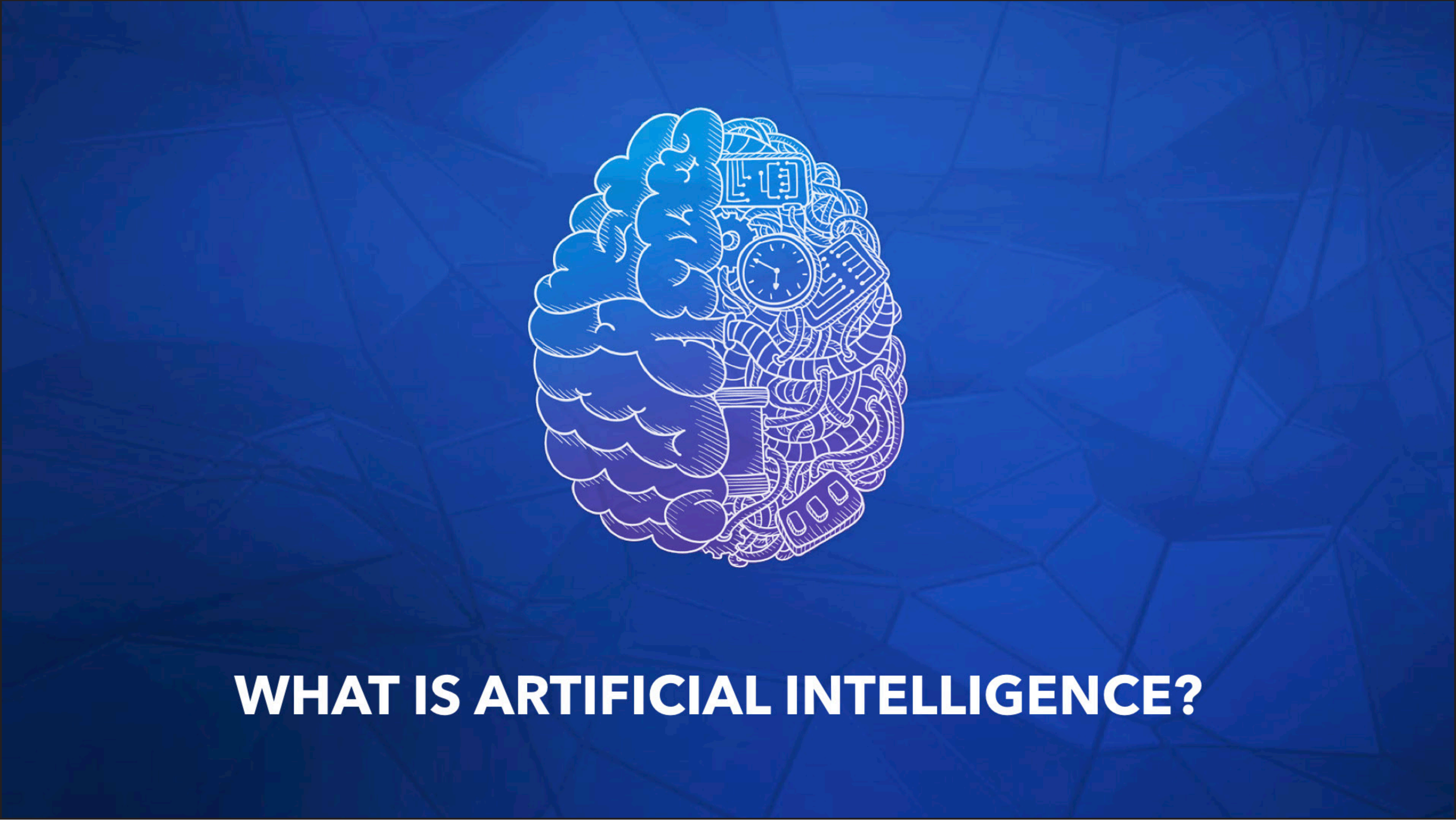
- To give the Post Secondary team a high level overview of these technologies for good general working knowledge



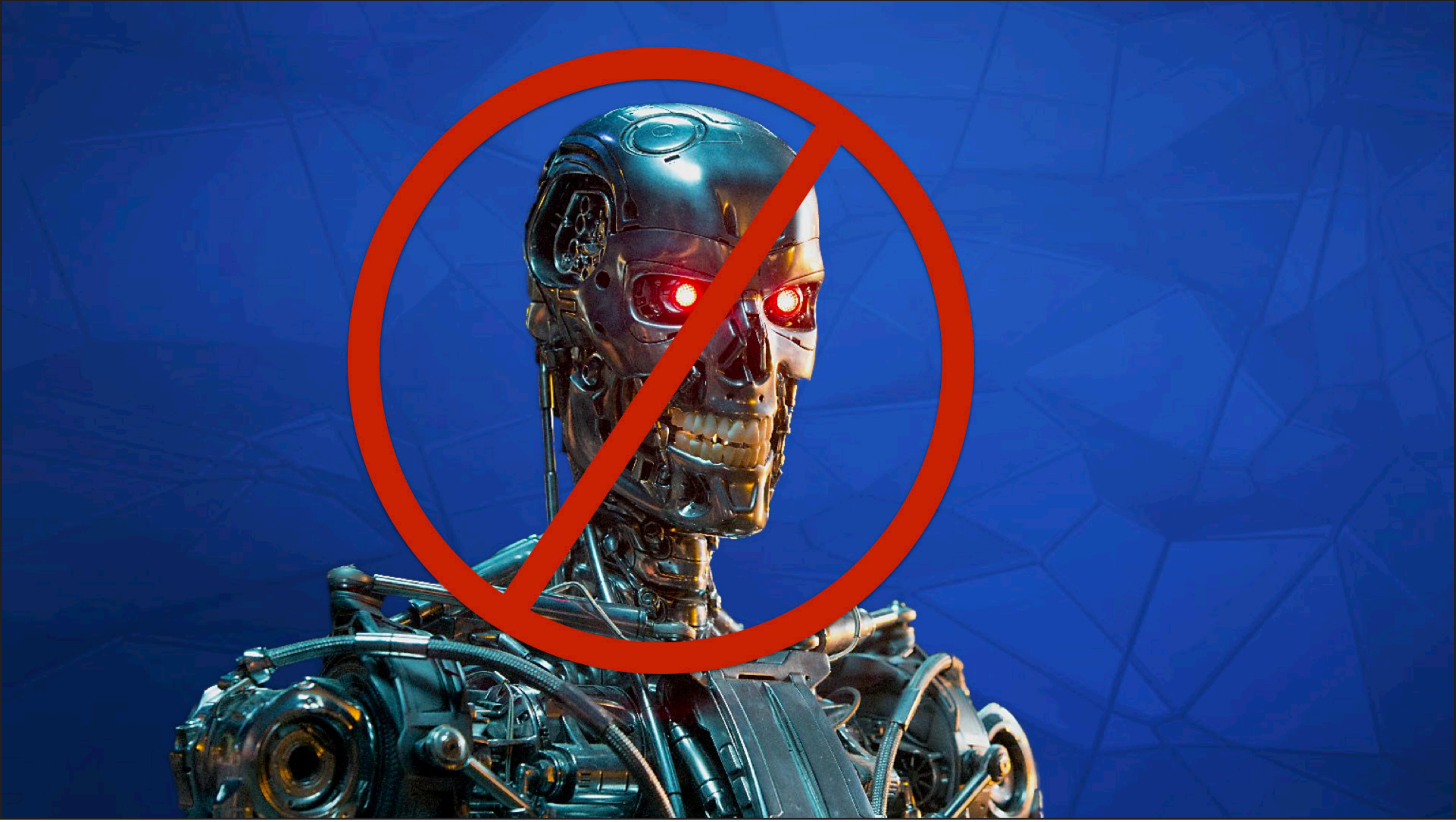
- 2. To share insights into how technologies in AI are leveraged today in Higher Education.



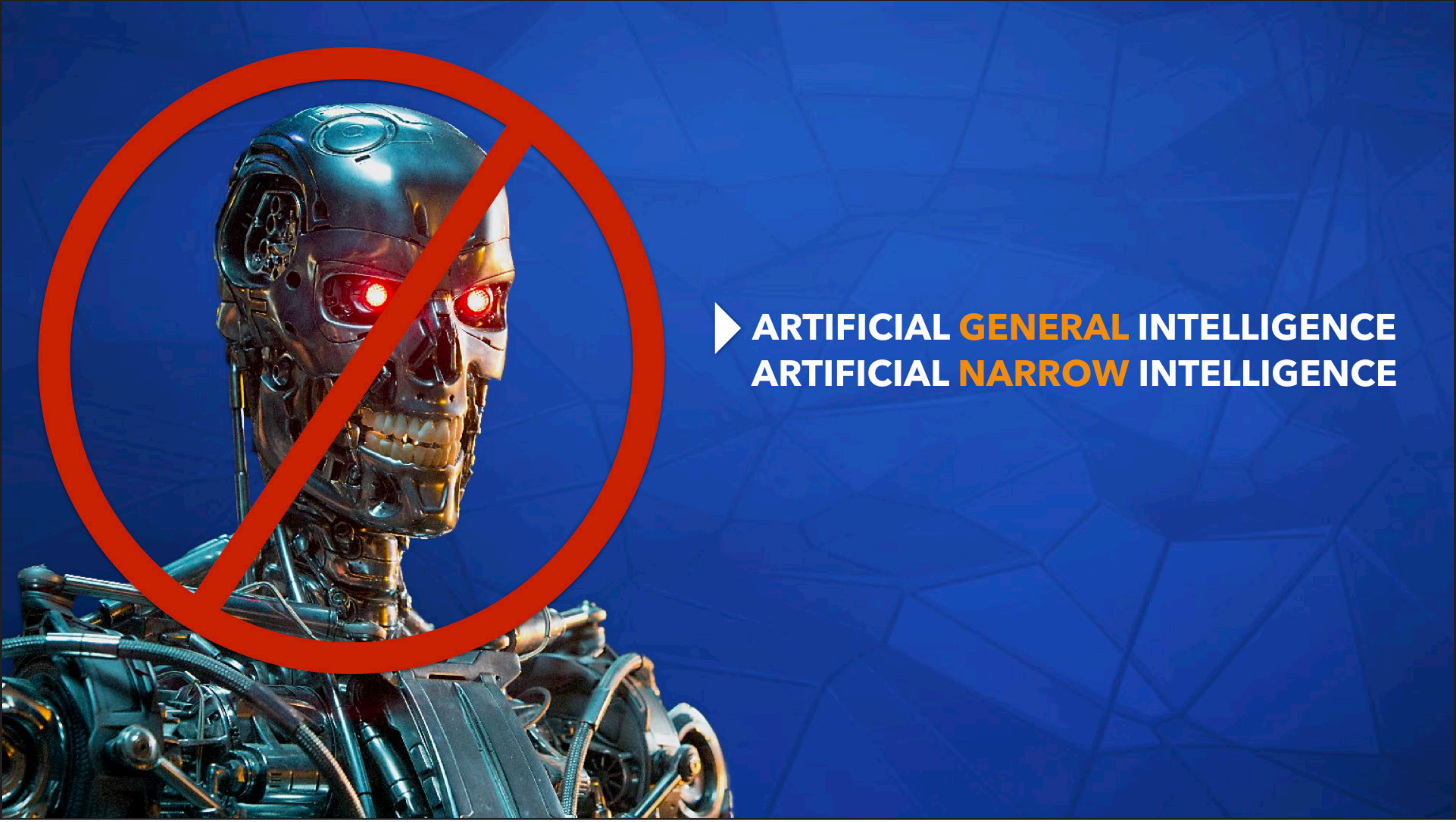
- 3. To evaluate opportunities to leverage these nascent technologies to meet Post Secondary's specific goals.



- So what is Artificial Intelligence?
- AI is an area of Computer Science that equips computers with cognitive functions which we normally associate with human intelligence.
- like learning and problem solving.
- It's an unassuming definition but of all the emerging technology topics we'll be covering, we find AI will be the most impactful, not just for education but for our lives in general.

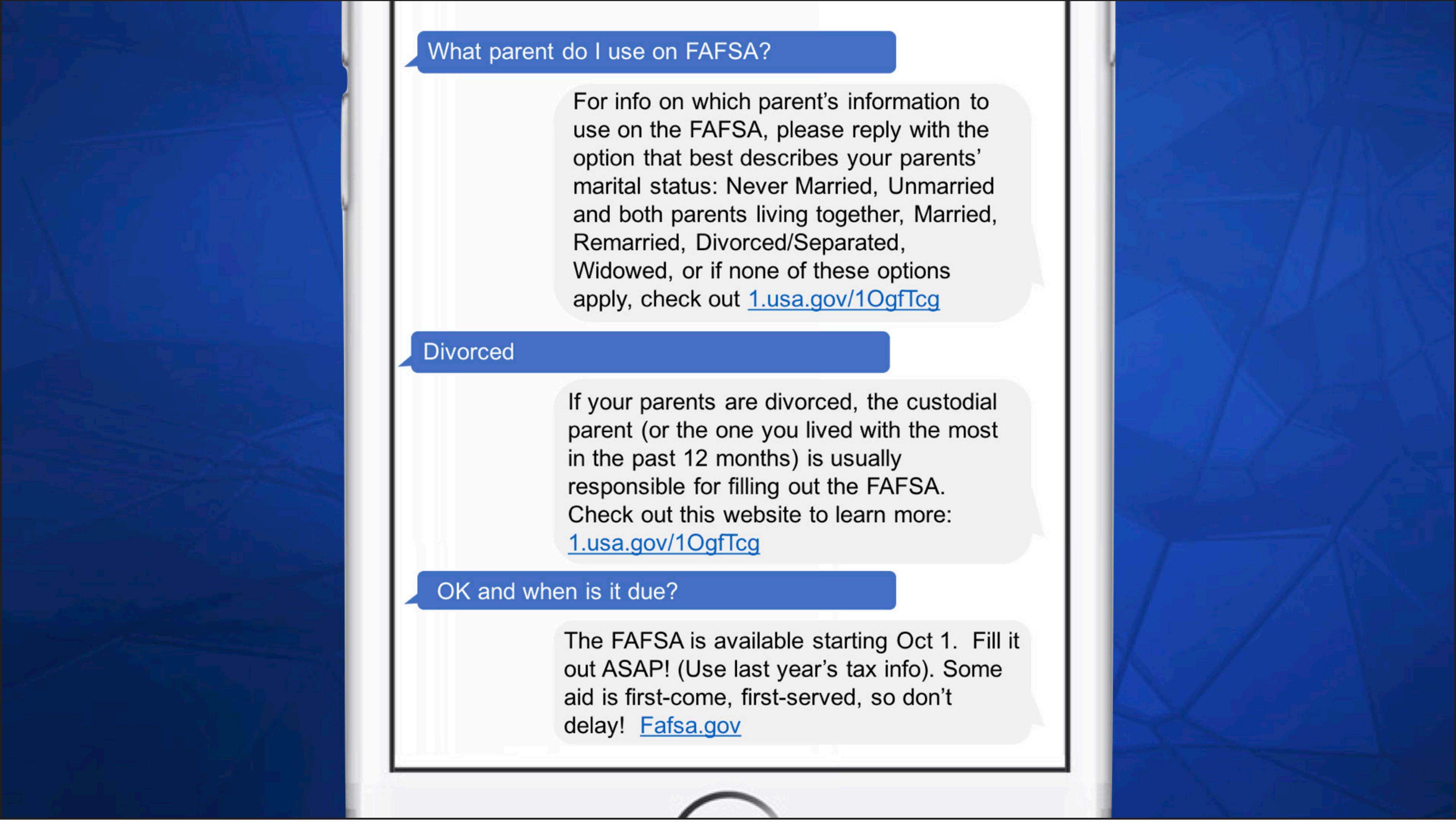


- and it's not because of those super intelligent robots from science fiction…

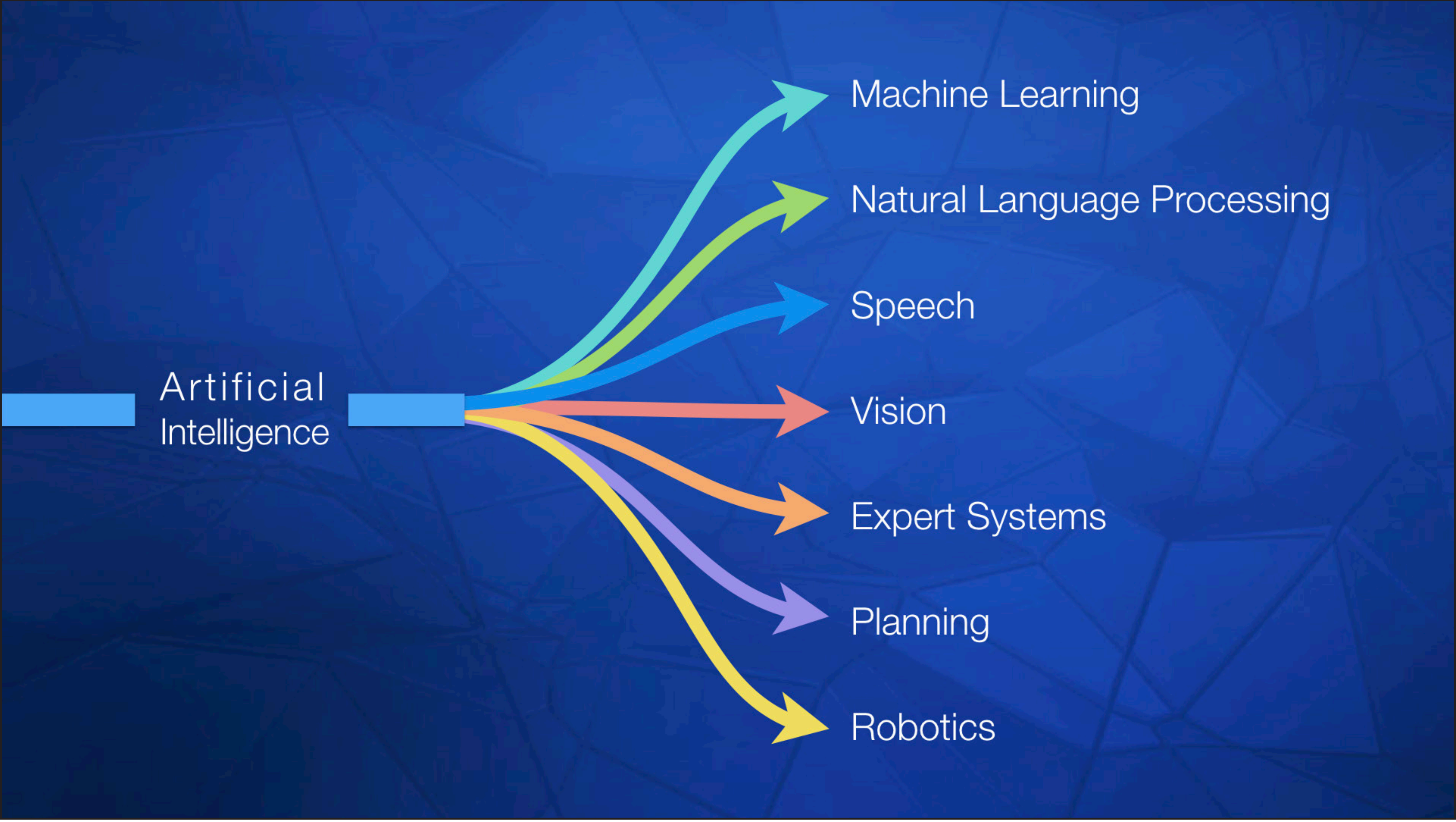


- Sentient machines actually fall under a branch of AI called Artificial General Intelligence, which is still more research and theory than anything else

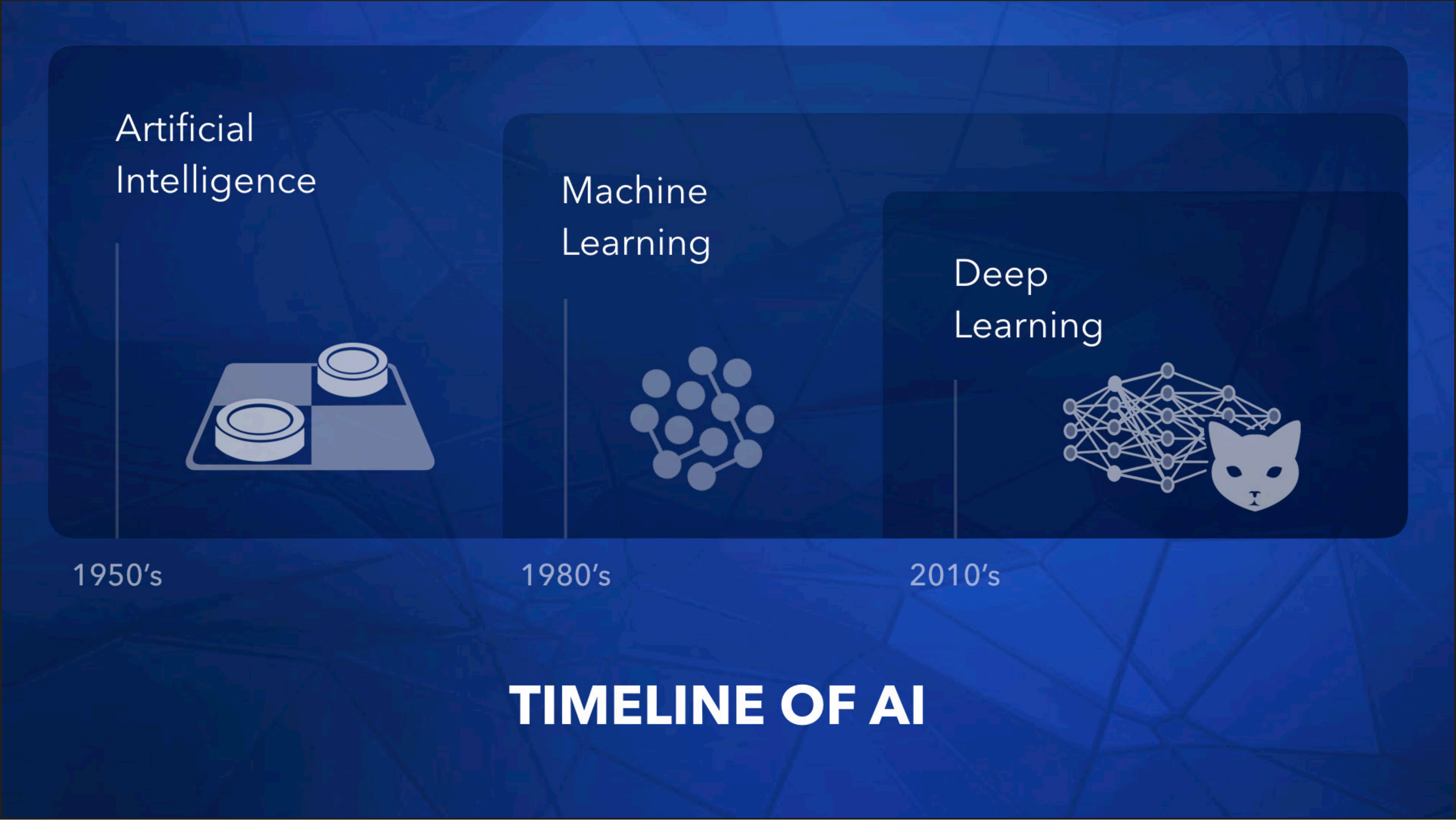
- The technologies discussed in this presentation and practically all AI technology today, fall under the the scope of Narrow AI.
- AI designed to solve singular problems or tasks, extremely well—often better than human specialists can, despite not actually having consciousness.



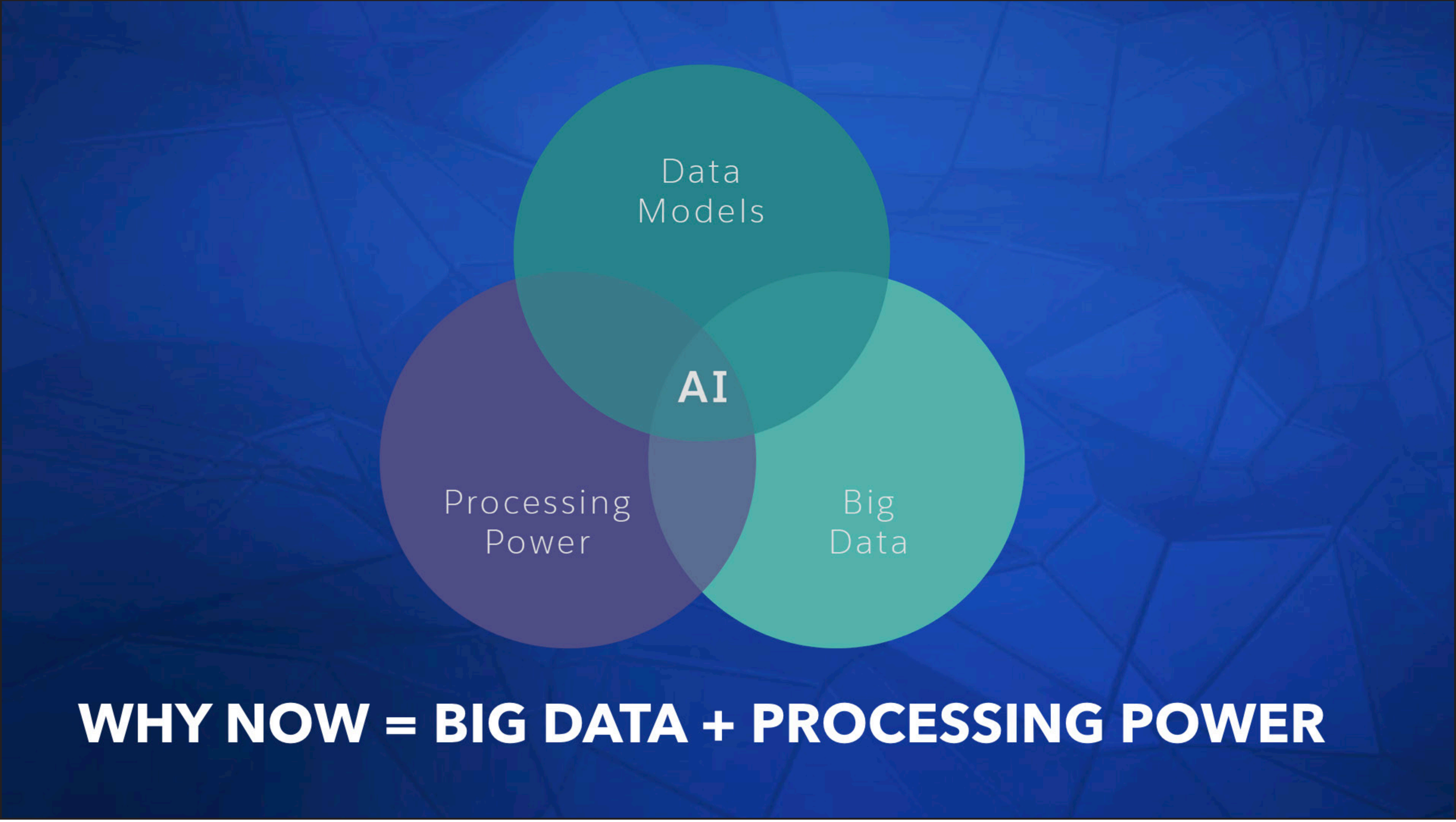
- For example this is a screenshot of a student texting an AI chatbot to help him fill out his FAFSA form for federal student aid.
- It runs 24 hours a day and can handle hundreds of students at once.
- This chatbot, named Pounce, uses a form of AI called Natural Language Processing to understand and response to requests.



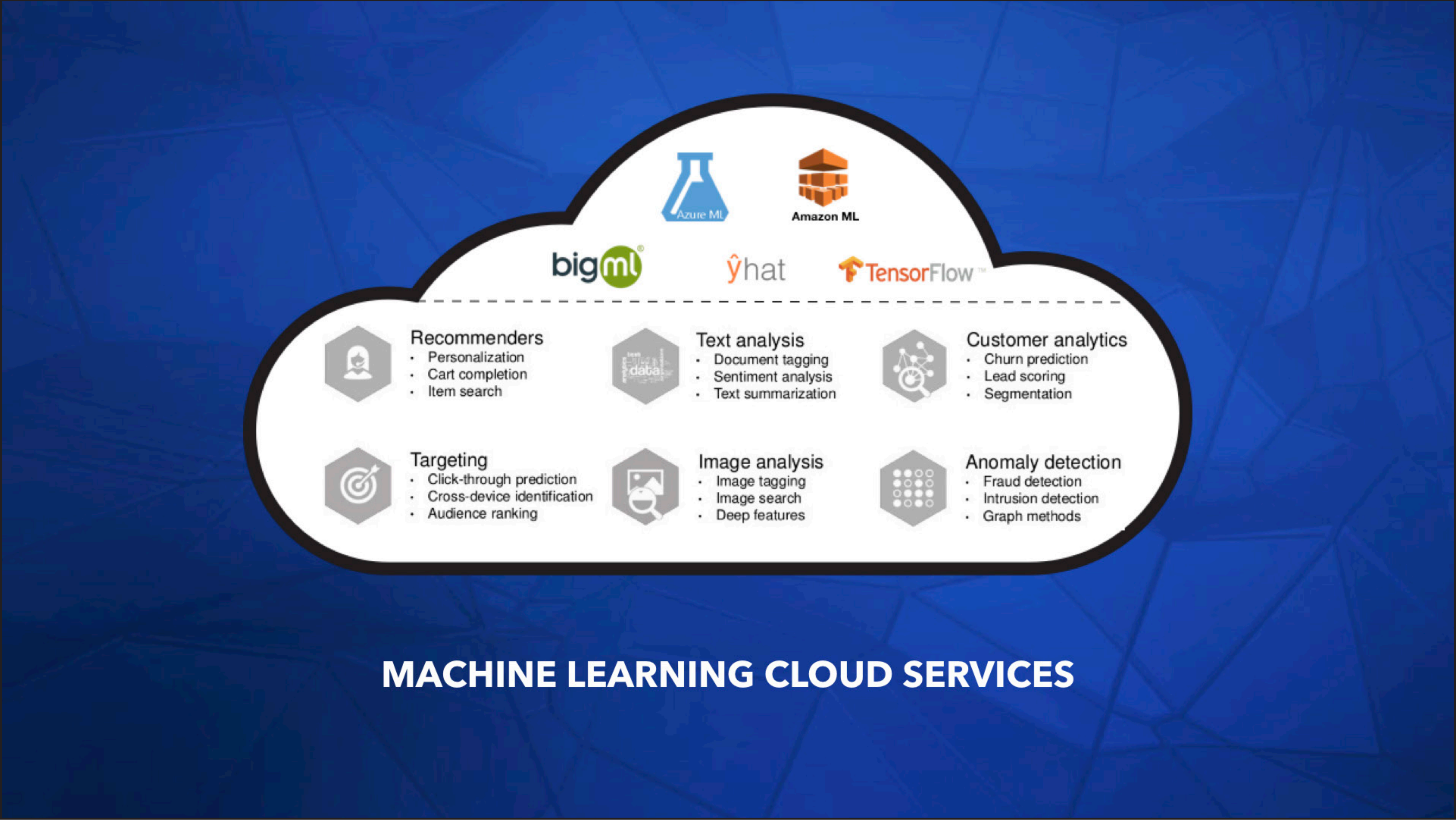
- NLP is one of several main branches of Artificial Intelligence from Robotics to Vision.
- We'll be focusing on just the branches with the most direct application to Higher Ed like Expert Systems and most importantly, Machine Learning.



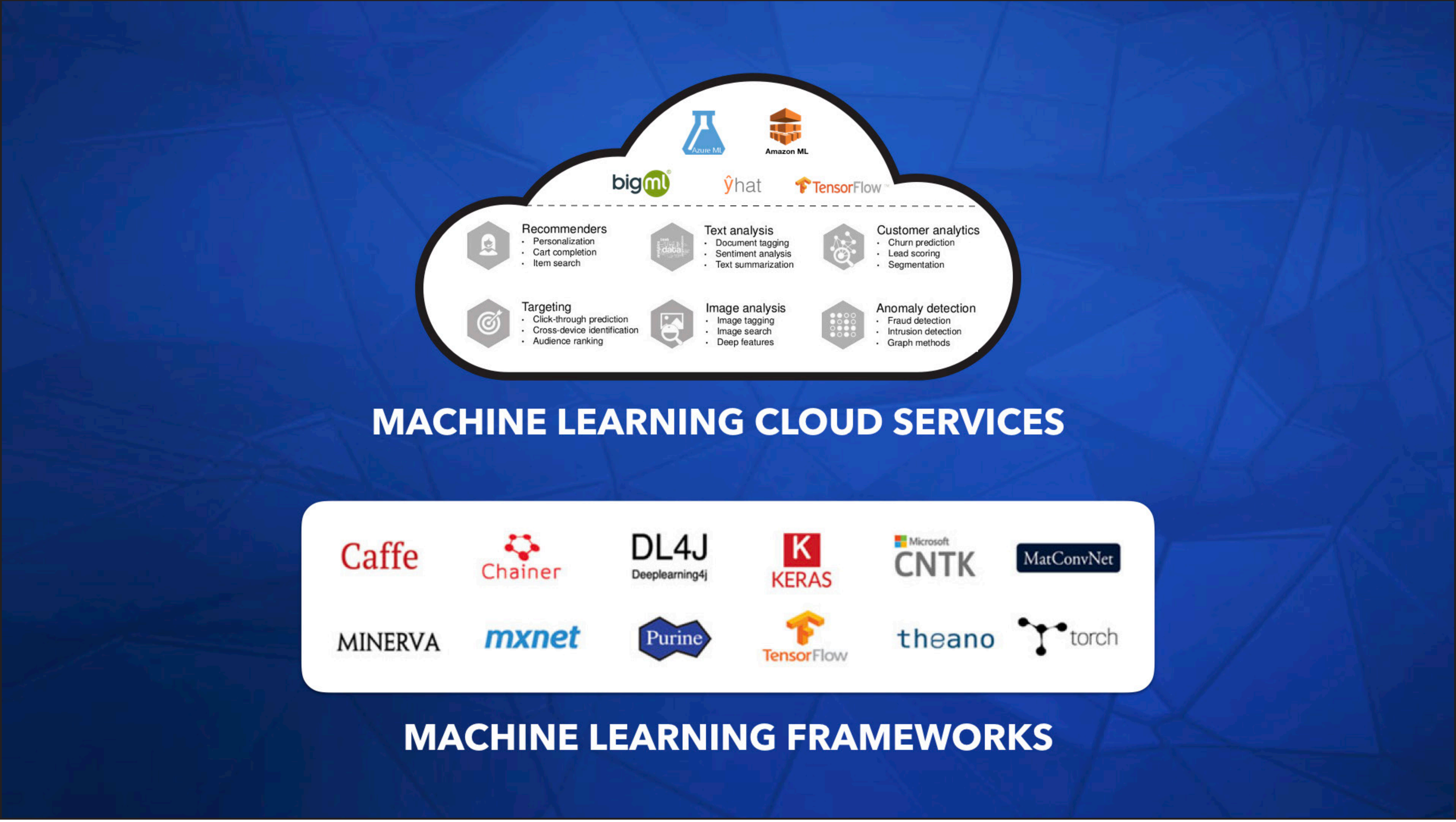
- Machine Learning plays such a significant role in AI that major periods of AI development are often defined by it.
- Long before we had the broad set of AI fields we have today, the concept of applying intelligence to computer tasks had it's start way back in the 50s
- by the 80's Machine Learning reorganized as its own sub-field and really started to flourish by utilizing methods and models borrowed from statistics and probability theory.
- The third and current period of AI only started around 2010 when a sub-field of Machine Learning called Deep Learning became practical.
- This field is not to be confused with Education's "Deeper Learning" pedagogy concepts, and it's a major driver of the the AI boom we're experiencing today.



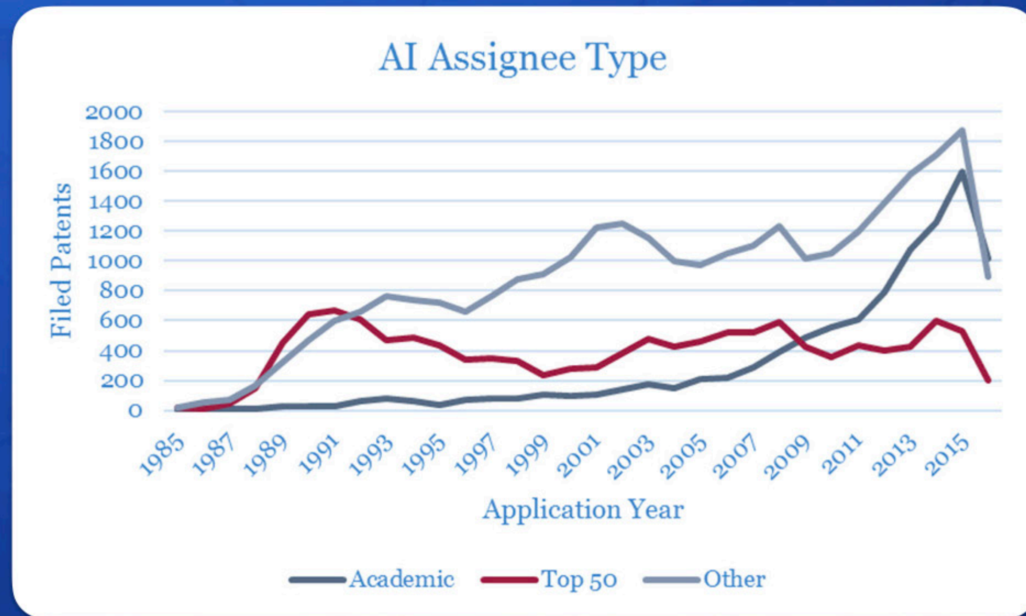
- 3 components were needed for this boom to happen:
- First we have the Data Model. Today's advanced algorithms, particularly in Deep Learning gives AI the ability to solve diverse problems too complex for human coders.
- Next we have Big Data. These new Data Models require massive amounts of data. Thanks to our modern digital infrastructure and cheap abundant data storage, we now have those large scale data sets.
- Finally Processing Power. Big Data isn't effective without an efficient way of processing it, and the parallel processing power of today's graphics cards and networked cloud computing make it all possible.



- Today even a regular individual can leverage these new Machine Learning technologies on the cloud, as cloud computing providers like Amazon Web Services and Microsoft Azure have integrated Machine Learning as a Service into their cloud offerings.

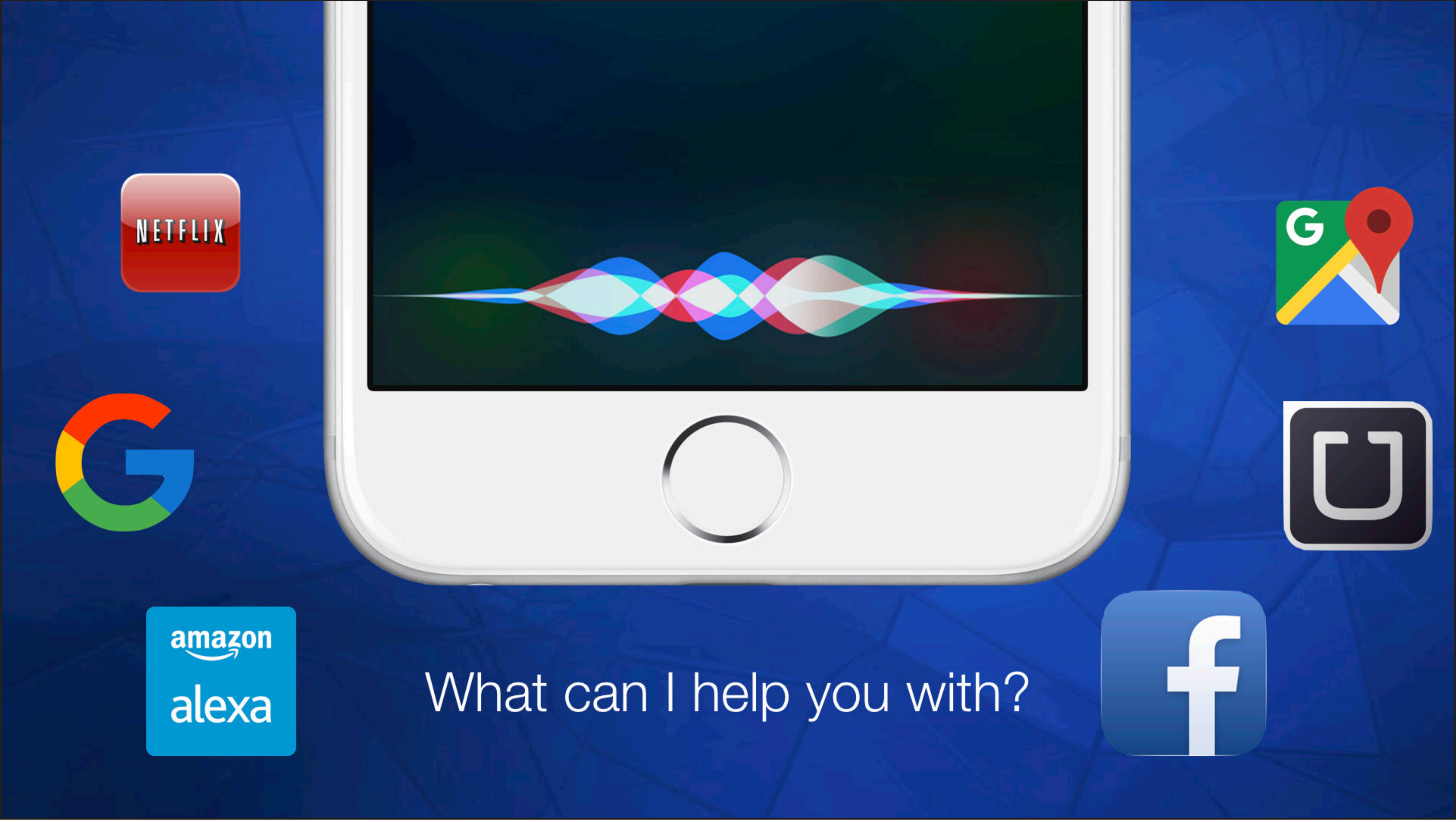


- And there's an ever expanding list of free and open source Machine Learning software as well.

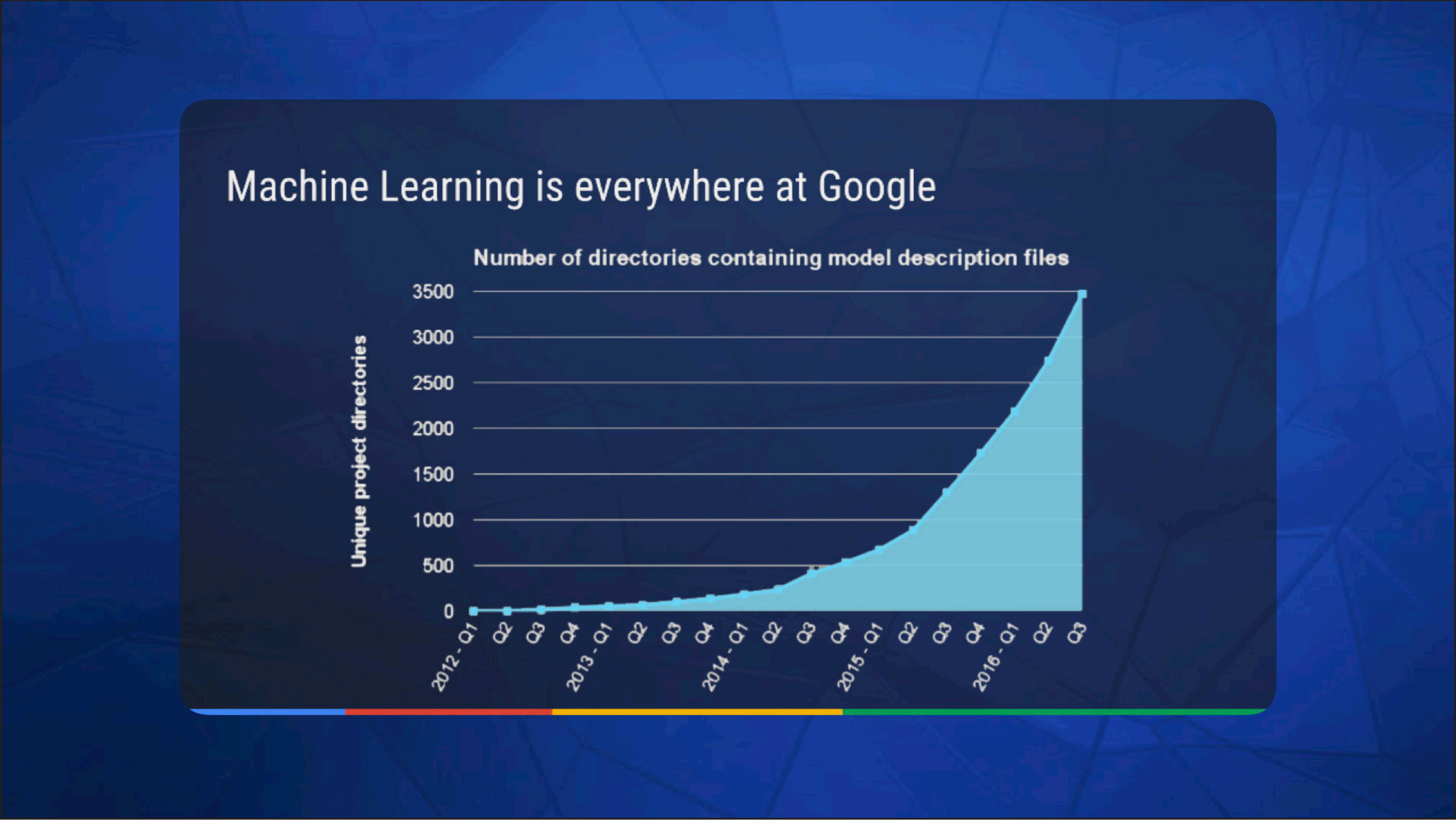


SOURCE: ClearViewIP Ltd.

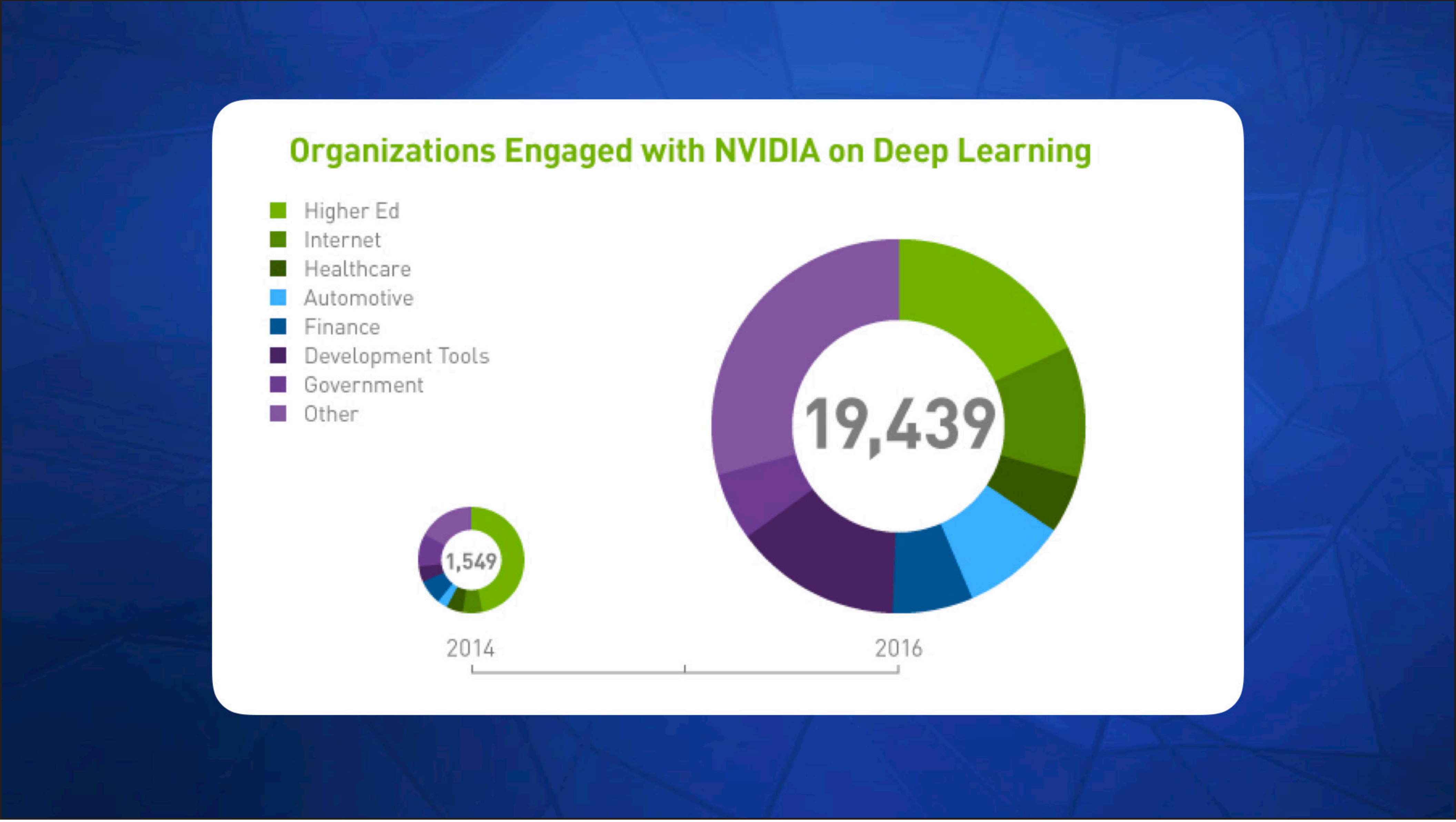
- Most of these open-source Machine Learning Frameworks have actually been released by pioneering institutions like Carnegie Mellon, NYU, and UC Berkley
- universities continue to spearhead much of this technology.
- On the right, we can see in dark blue that 40% of AI patent assignees in 2016 are coming from academic sources.



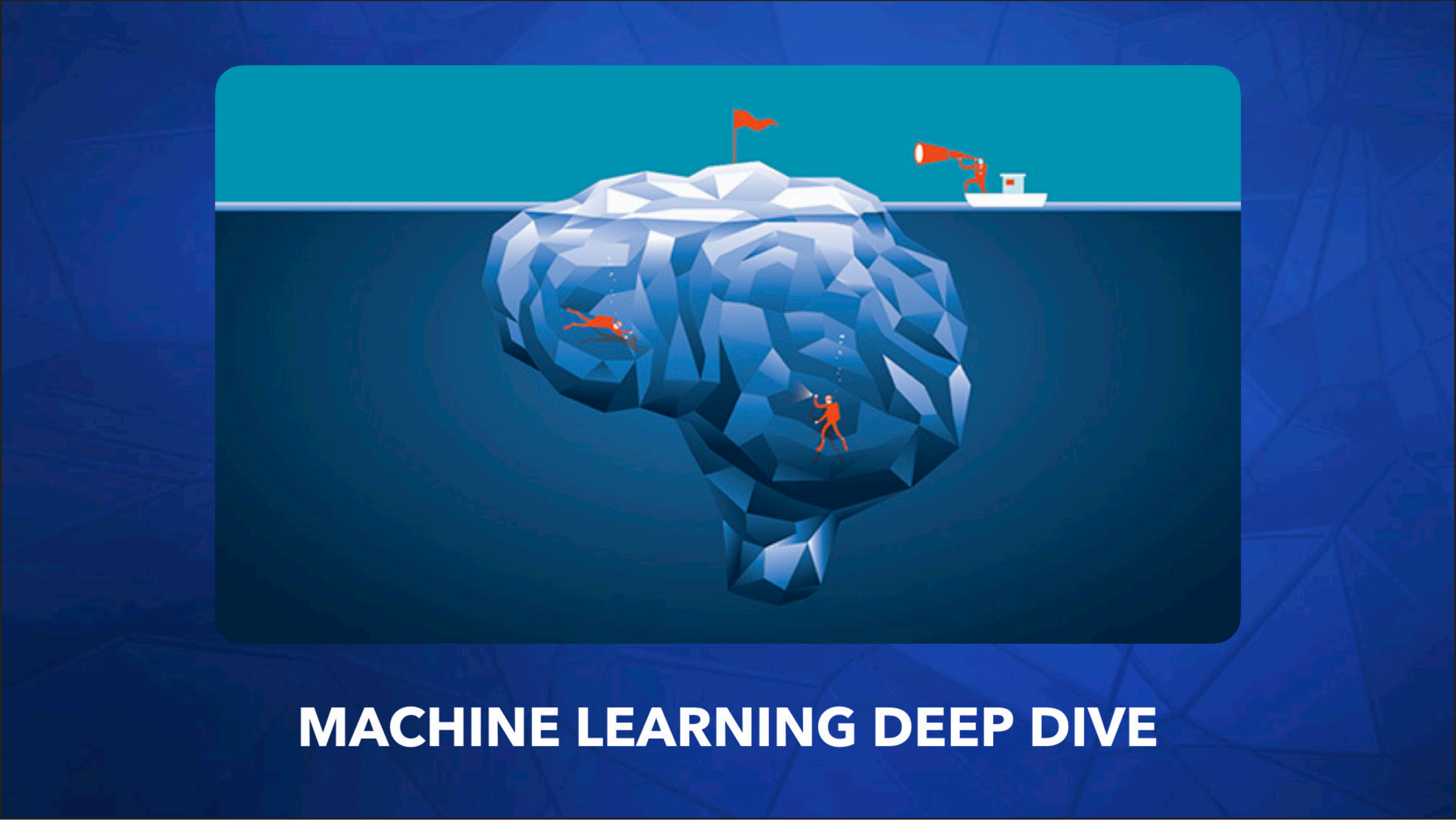
- Today AI technology can be found all around us.
- For example, If you've ever used voice assistants like Siri or Alexa, these leverage Natural Language Processing and Deep Learning for natural sounding voices
- If you've ever had Facebook automatically suggest a photo tag, that's using DeepFace, Facebook's deep learning facial recognition system.
- And if you've ever done a Google search—it's been utilizing a machine-learning AI system called RankBrain since 2015.



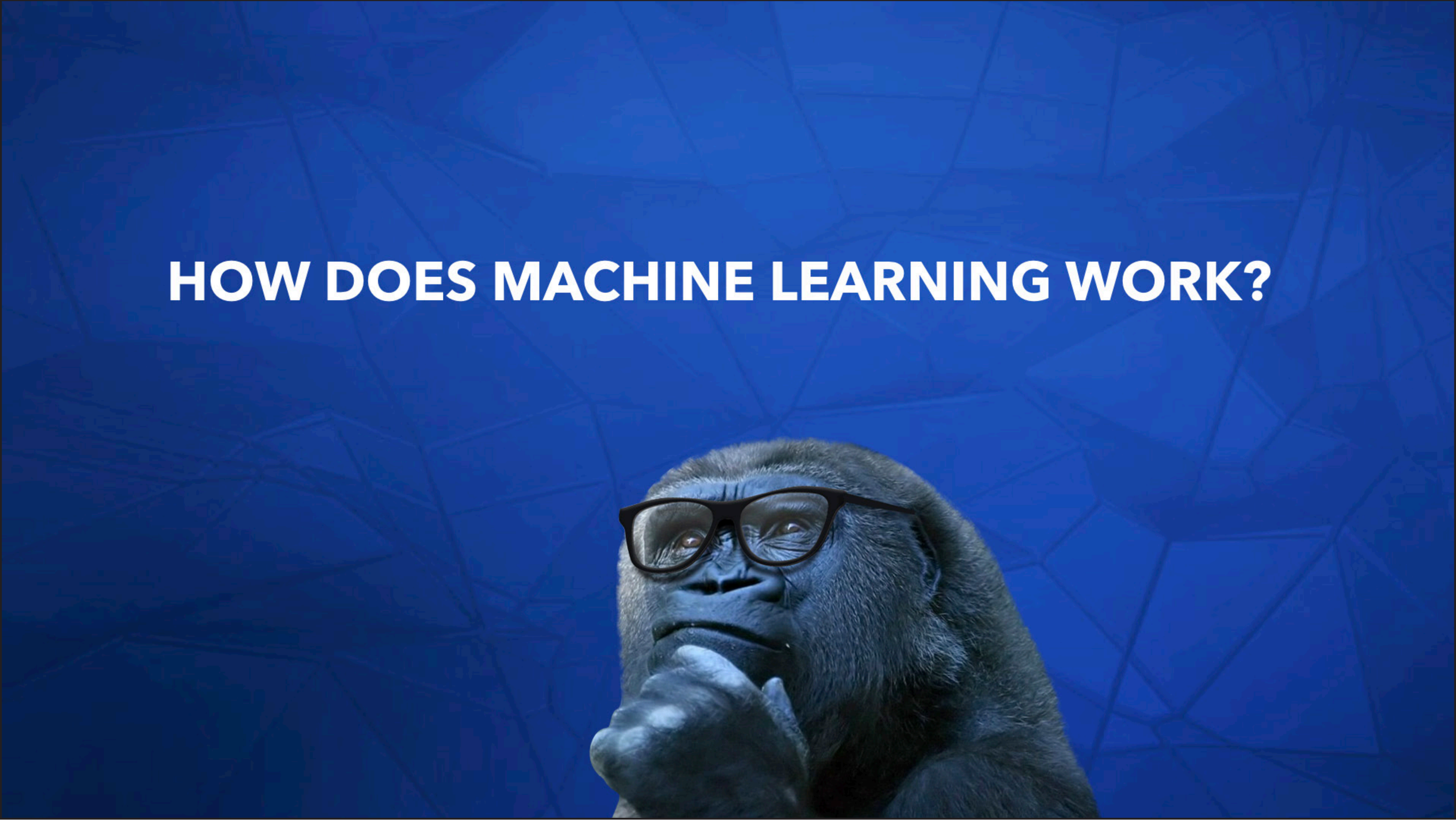
- Google in particular has been a big proponent of AI, and has over 4,000 projects harnessing machine learning.
- You can see that back in 2012, only five years ago, they had zero.
- Now it's leveraged in most Google products from Android and Chrome, to Maps and Cardboard.



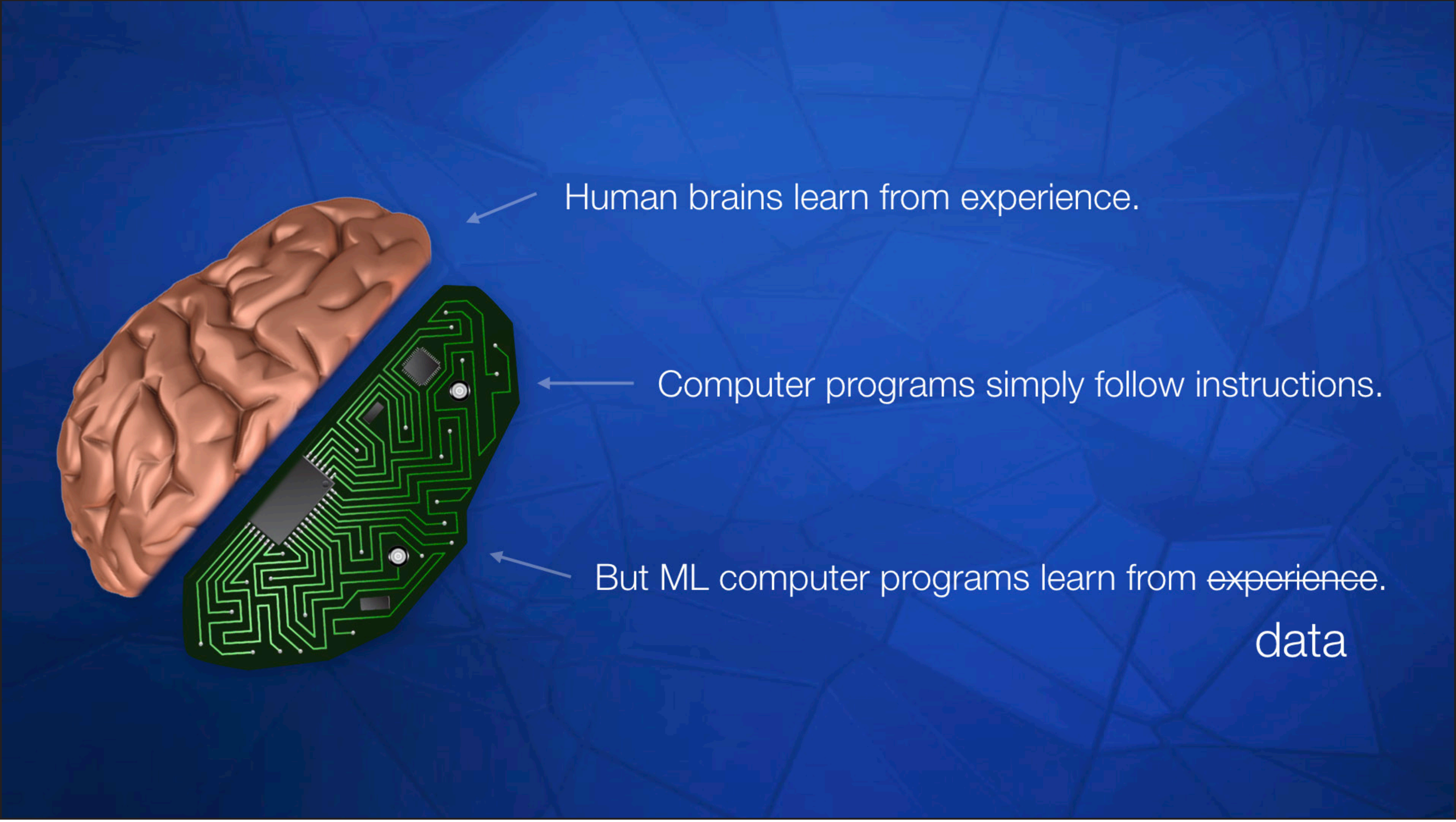
- Another pioneer in machine learning is Nvidia who produces graphics cards utilized heavily in this field.
- They've noted explosive interest across industries, with Higher Ed notably on that list.



- Now let's take a closer look at that big important piece of AI, Machine Learning and it's offspring, Deep Learning.



- Machine Learning is often thought of as a mysterious black box full of answers.
- But you don't need a technical background to understand how it works.
- Getting familiar with the conceptual parts should be helpful in assessing it's use in Higher Education.
- And we won't need to get into any esoteric computational details to do that.



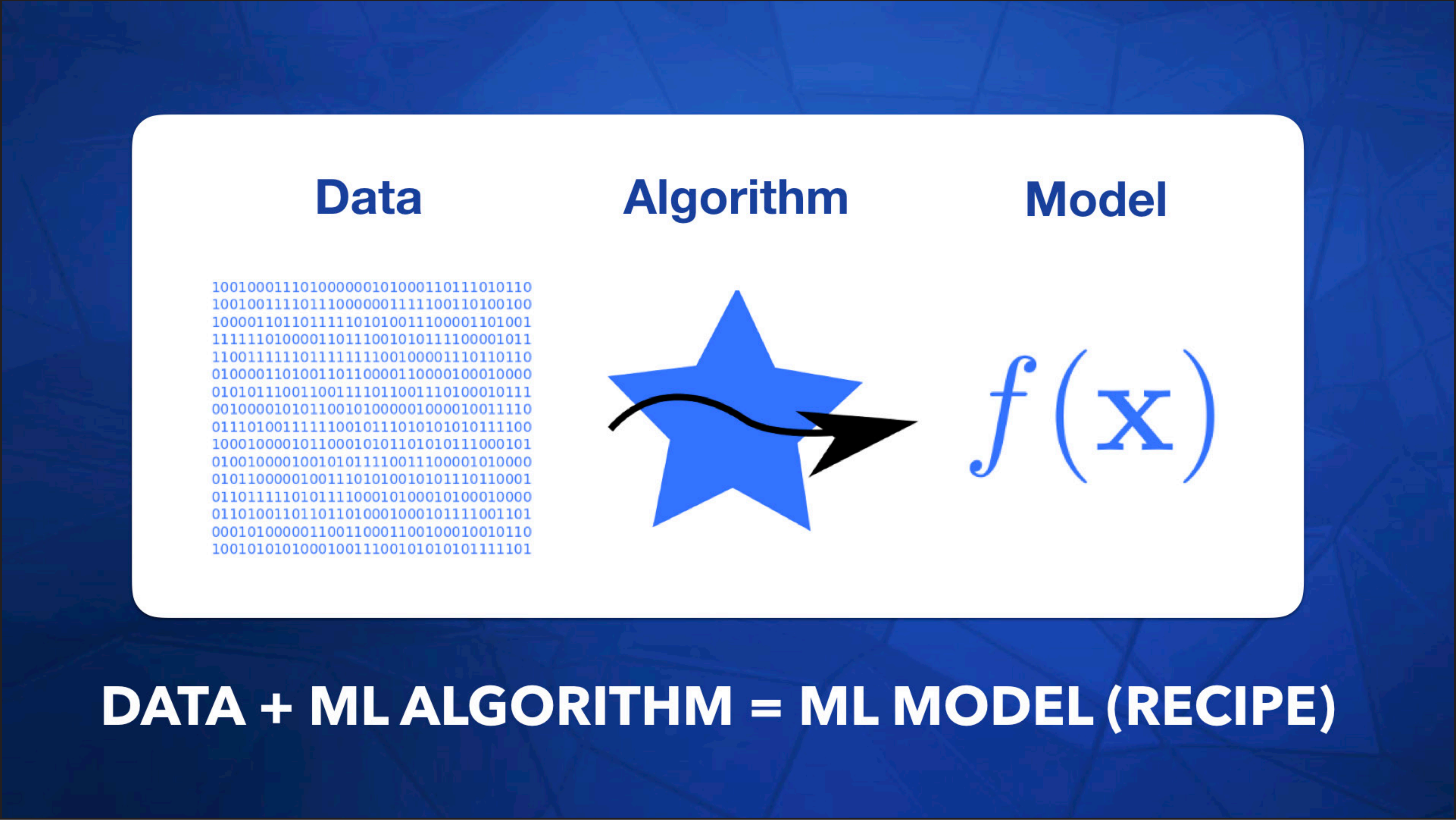
- From a theory perspective, a good comparison is to how human's learn.
- Traditional computer programs simply follow coded instructions like an advanced calculator. If X, then do Y. While human brains learn from experience.
- But with Machine Learning, computer programs learn from experience too. That experience is called "data".

- Lets look at an example. Let’s say we wanted to create a Machine Learning formula that detects when students are at risk of dropping out, a common use case for retention solutions.
- If we had access to enough historical institutional data for past students, pulled from their SIS, LMS, transcripts, and related sources
- we could test that data through the right Machine Learning Algorithms until we found a valid set of predictors to accurately forecast future at-risk students.

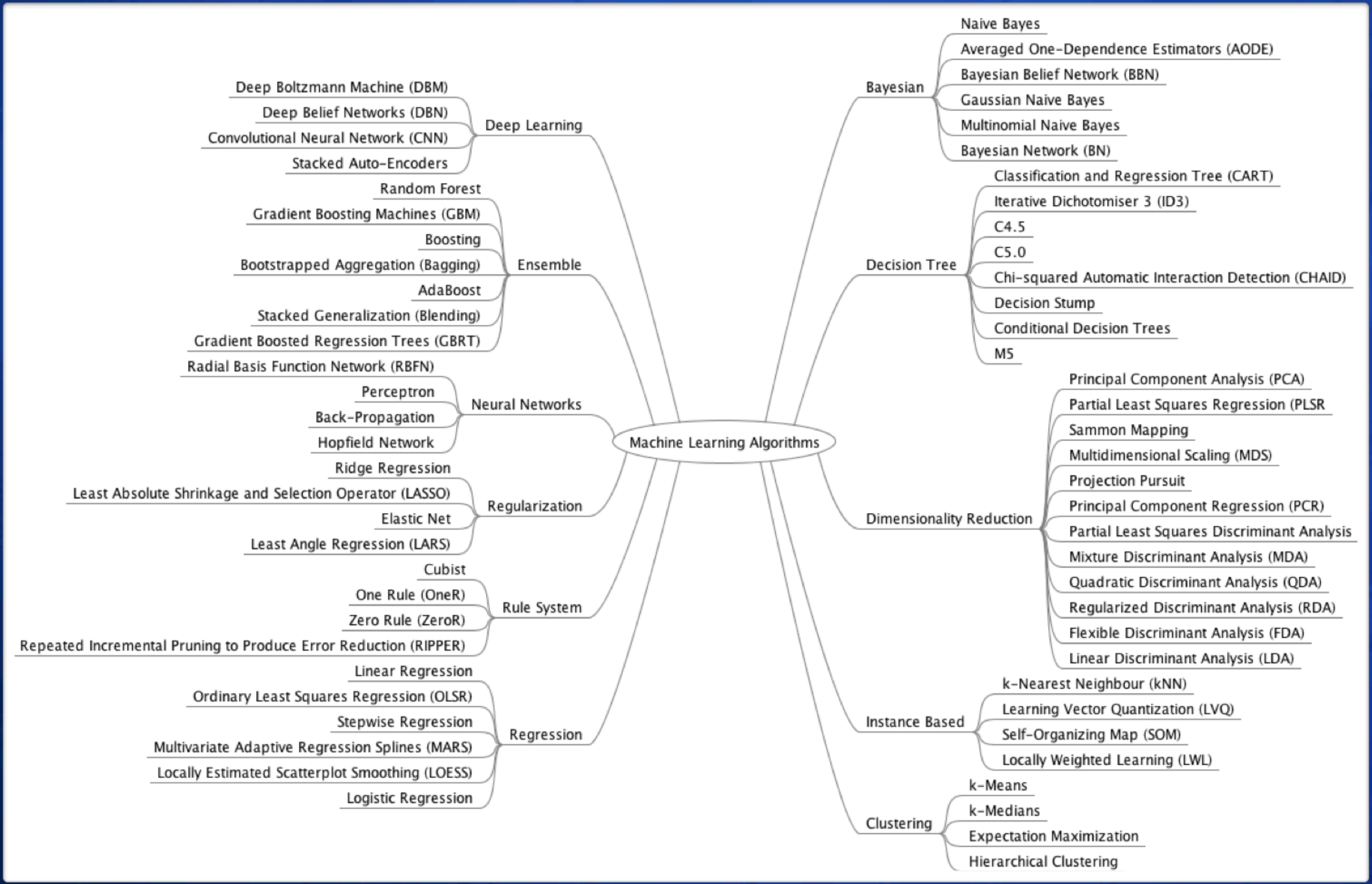
- The formula Machine Learning generates for us could find the common denominator(s) between combination of courses, missed days, excessive credit hours, financial issues, or even a particular dorm hall to be determining factors of at risk students.

MACHINE LEARNING
VS
PREDICTIVE ANALYTICS

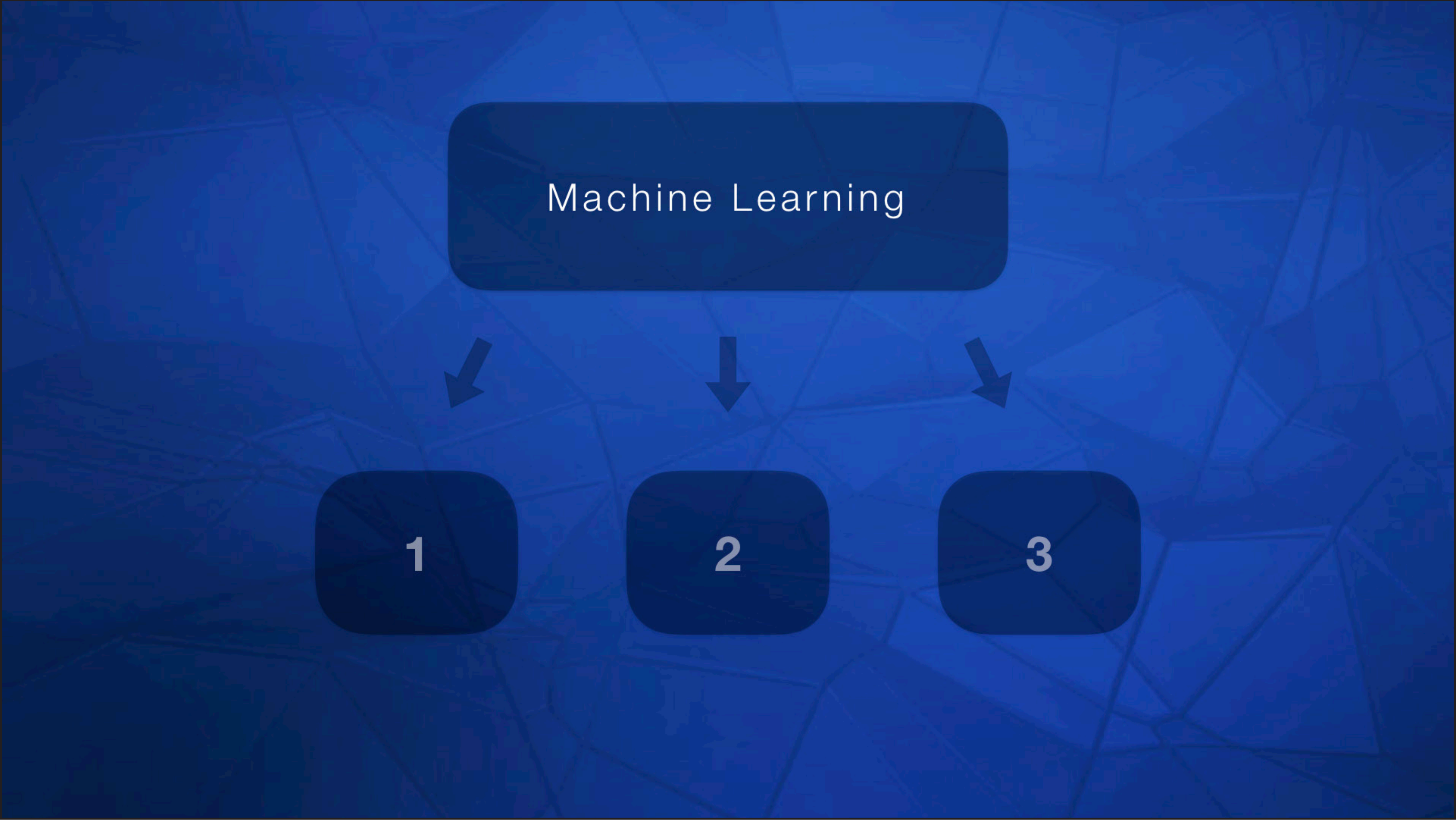
- You might ask, wait, doesn't that sounds a lot like Predictive Analytics: Using historical statistical data to predict future outcomes?
- And you'd be right, Machine learning uses statistical algorithms to make high end predictions on data just like other forms of Predictive Analytics.
- Machine Learning however, has much broader use cases than just Predictive Analytics, and when used in this way it can be considered a pioneering technology in predictive analytics with a host of advantages.



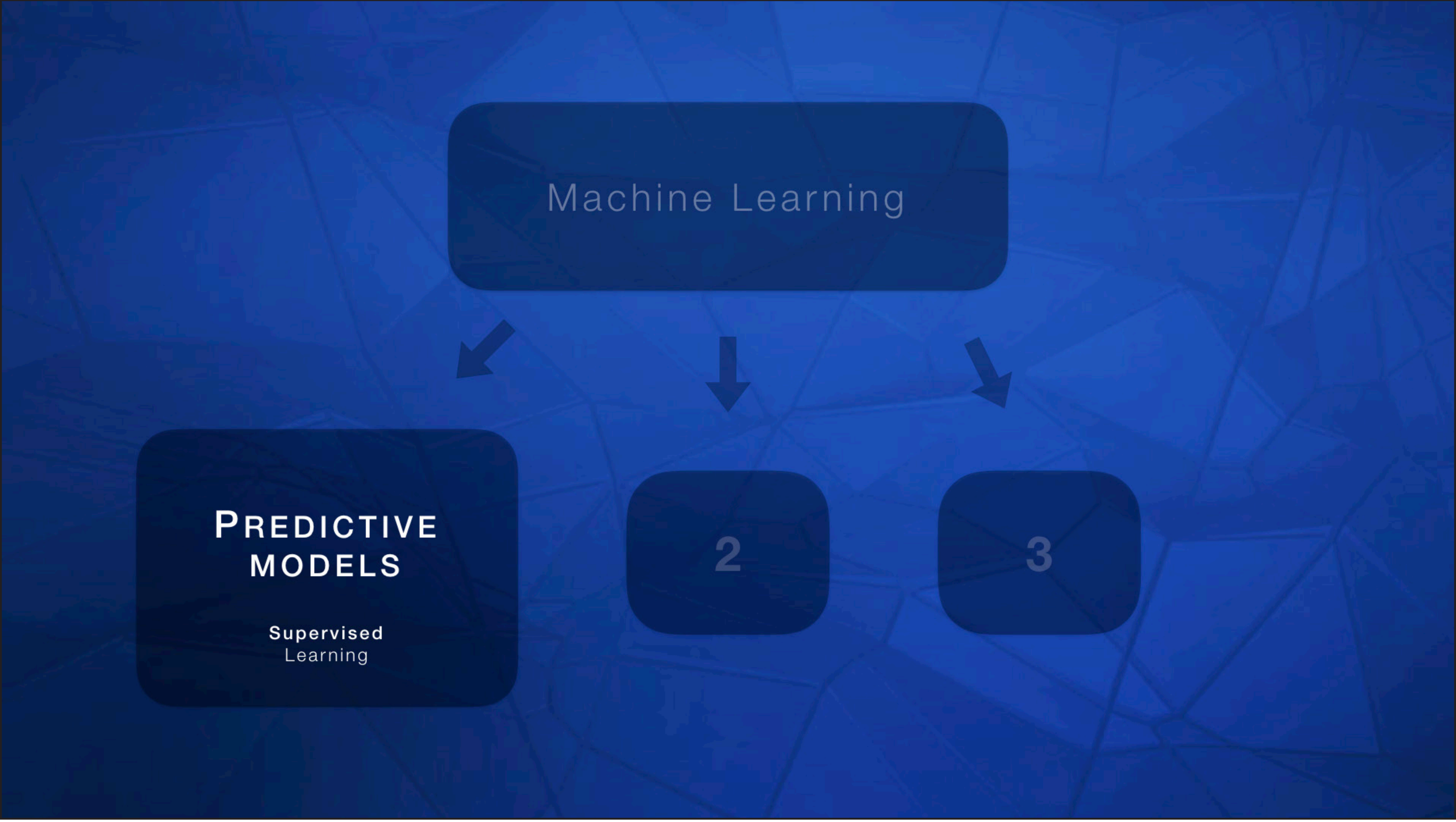
- Most significantly, Machine Learning is all about generating accurate, re-usable, predictive formulas, automatically, through data training.
- In this case our institutional records are passed through a specific machine learning algorithm (picked by data scientists) which generates a predictive formula unique to the patterns and cases in our data.
- That solution is automatically written down into a self-contained formula called a Machine Learning Model.



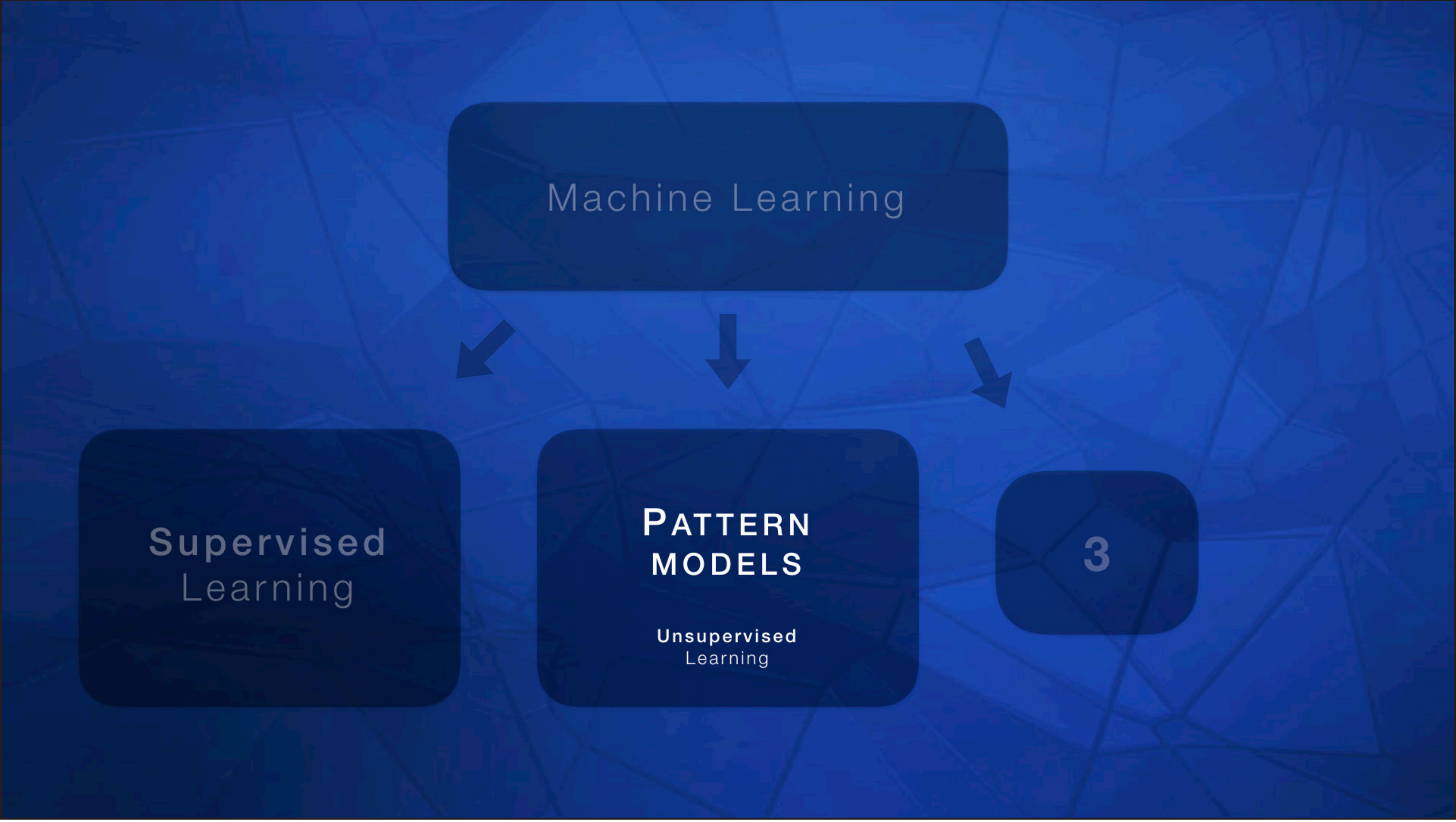
- And Data scientists don't have to rely on making these predictive formulas from scratch, that workflow is streamlined too.
- They tap into an extensive field of well documented task-based Machine Learning Algorithms organized by type of solution.
- These are the generic formulas data scientists train test data against, to automatically generate our data-trained, Machine Learning Model formulas.



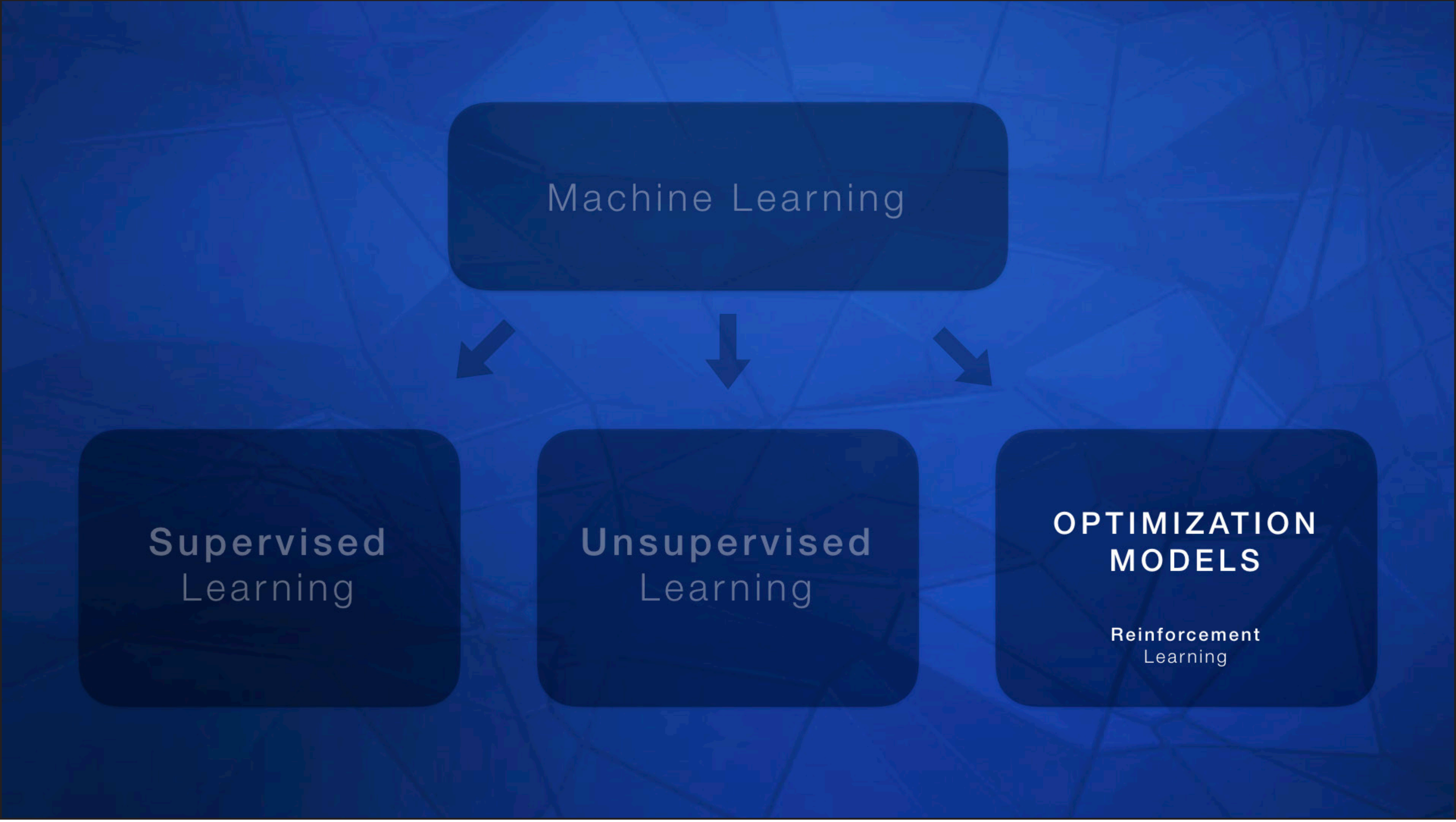
- The Algorithms available to Machine Learning fall under 3 major types of Machine Learning. These generally provide solutions to 3 different types of problems.
- Here they are with our own theoretical examples of Student Success modeling, just to get you thinking about the types of solutions possible.



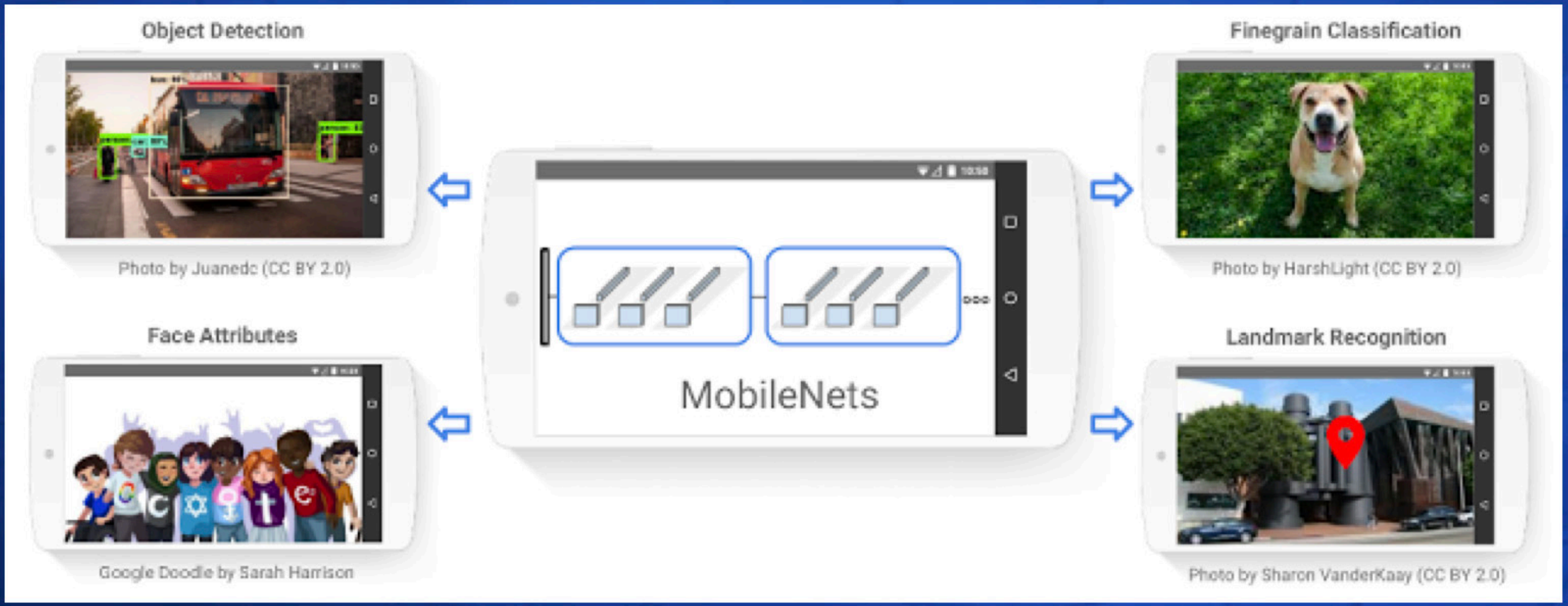
- The first, Supervised learning uses algorithms to make predictive models based on known past outcomes. This is the most common type of machine learning.
- Example case: To create predictive models that use historic student enrollment data that include graduation status to predict at-risk new enrollments.



- In Unsupervised learning, algorithms don't utilize previous outcomes but instead find hidden patterns, groupings, and anomalies in your data.
- Example case: To create models that identify common pitfall patterns that lead to student failure, or models to identify and group students by engagement strategy to optimize retention.

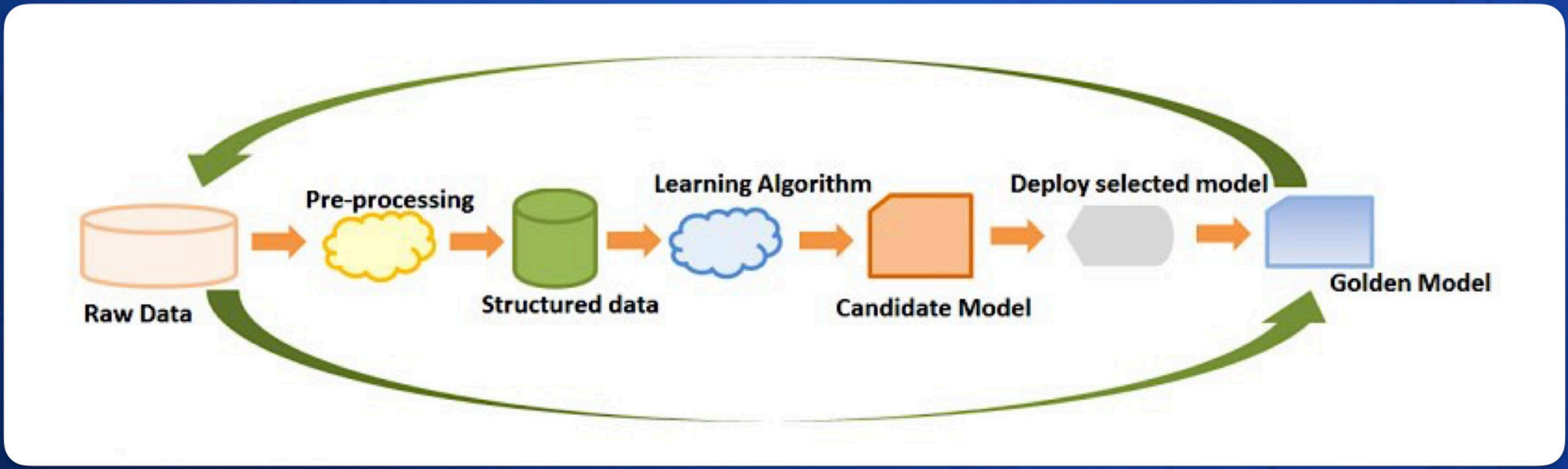


- Reinforcement learning: is inspired by behavioral psychology. It's used to find optimal paths for achieving goals by using reinforcement. Kind of like that children's game Hot-or-Cold.
- Example case: to create Models powering Course or degree navigation systems



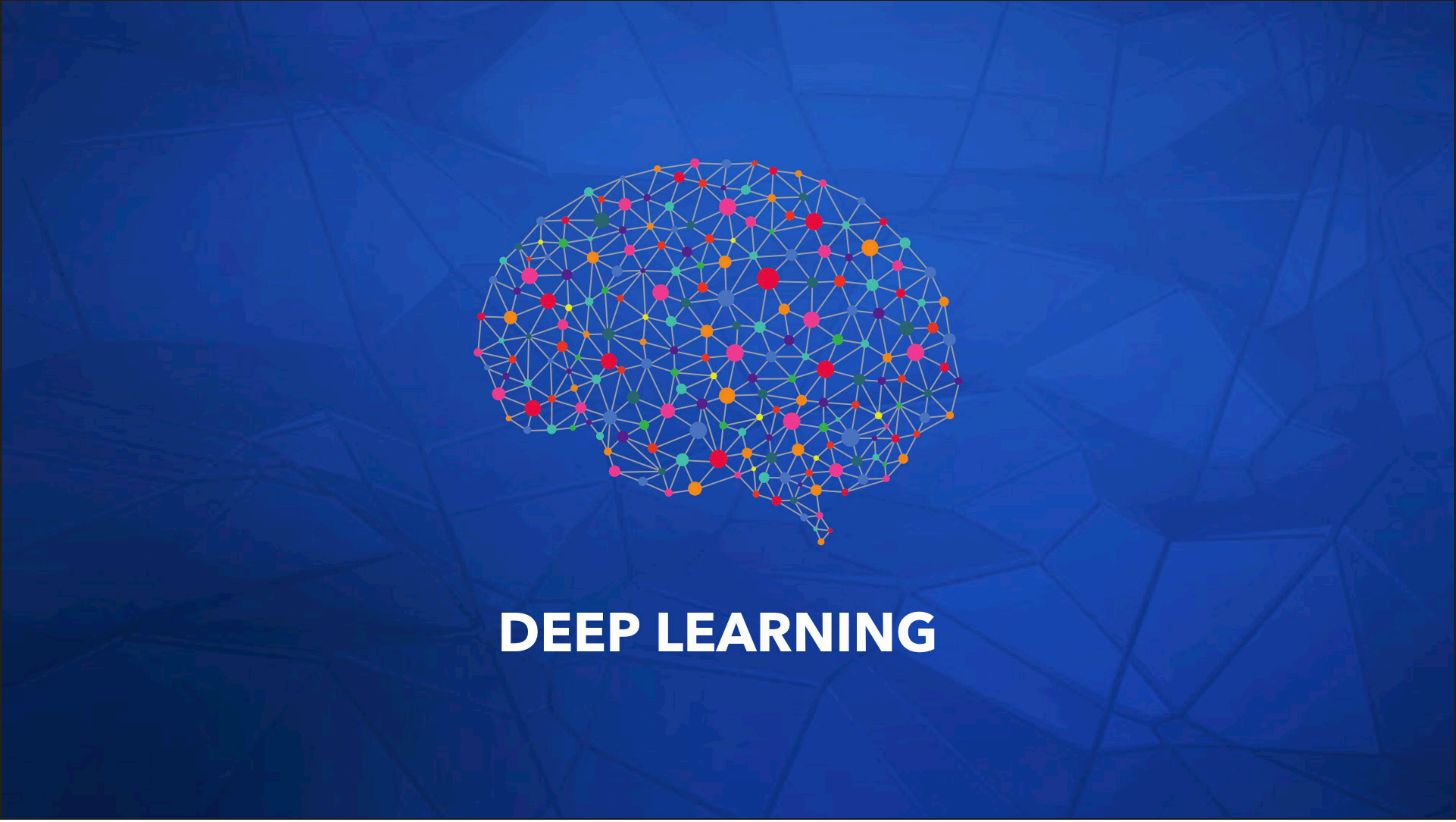
POWERFUL STANDALONE PREDICTIONS

- Once created these Machine Learning Models behave like self-contained logic components within an application.
- This is a pretty powerful concept. This reusable formula can be duplicated, distributed, and deployed in numerous applications including on mobile,
- making accurate predictions against new data without having to be tied to data warehouses.
- Google demonstrated the power of this when they released MobileNets this year, a series of mobile optimized vision models that allow applications to detect objects, faces, and features in photos, even without an internet connection.
- This is something that previously required cloud-based APIs or a powerful machine to do.

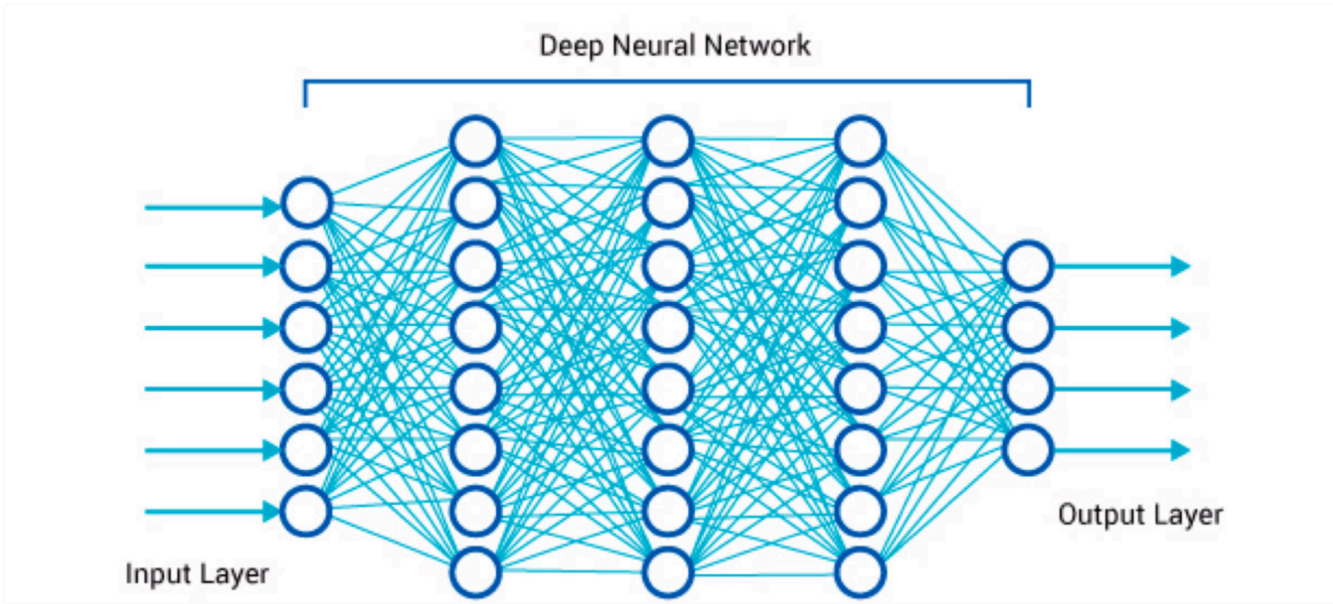


LIFECYCLE OF A ML MODEL

- The lifecycle for a Machine Learning Model includes 3 phases: data processing, model building, then deployment and monitoring.
- Step 1: A "raw" data set, is pre-processed into "structured data"—this just involves formatting it into labelled columns, similar to a spreadsheet.
- Step 2: This uniform structured data is then processed through a Machine Learning algorithm in a step called "training". This "training" process usually requires thousands, or sometimes millions, of quality data entries to generate an accurate predictive formula for future data.
- Step 3: That accurate predictive formula is our "Golden Model", which is then integrated into the final application to be leveraged with new data. This model can be further updated and tweaked over time as algorithms are improved, and more training data is acquired.

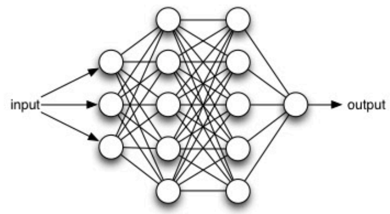


- Ok we've discussed how the magic of Machine Learning works, now what about it's offspring Deep Learning?
- As we mentioned Deep Learning is a sub-set of Machine Learning that has really taken off in the last few years.

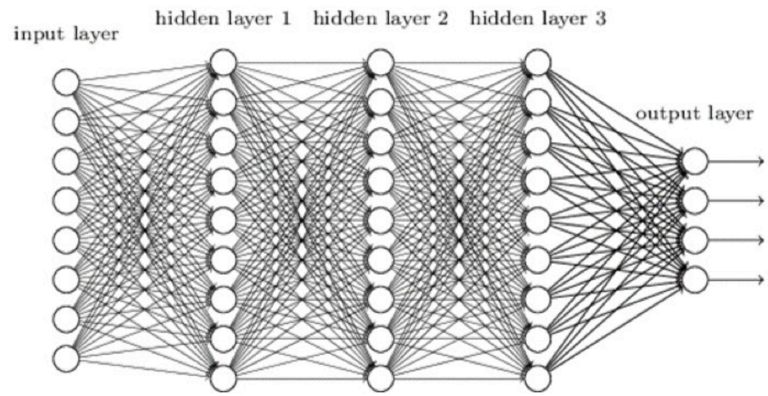


DEEP LEARNING = NEURAL NETWORKS

- Deep Learning was originally inspired by the way neurons work in the brain so it’s methods are aptly named “Neural Networks”.



Simple Neural Network

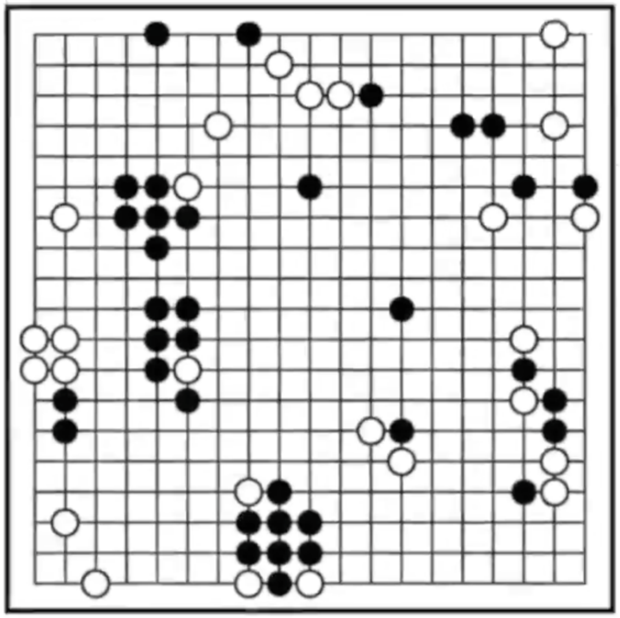


Deep Neural Network

DEEP NEURAL NETWORKS

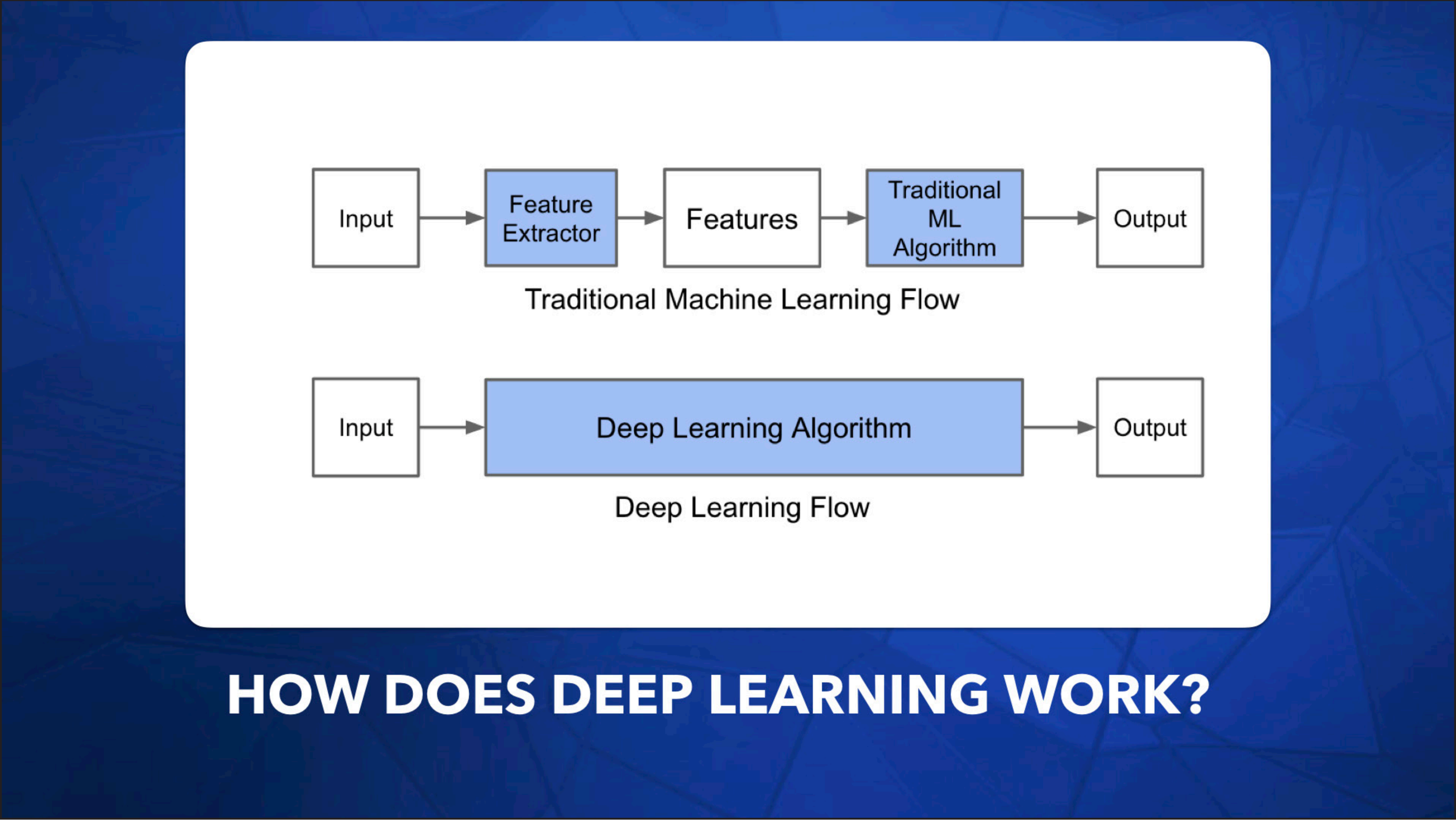
- This was actually an old AI concept but modern techniques moved it out of the realm of novelty when they significantly scaled up the network layers and the amount of data pushed through them, hence the overarching name “Deep” Learning.
- That scaling up turned out to be a major break through enabling many of the innovative applications of Machine Learning we see today like natural voice assistance and self driving cars.

- Google's AlphaGo recently showed off the potential of Deep Learning when achieving a goal that many experts predicted was still a decade or more away.
- This year it beat the human world champion at the ancient Chinese board game Go…

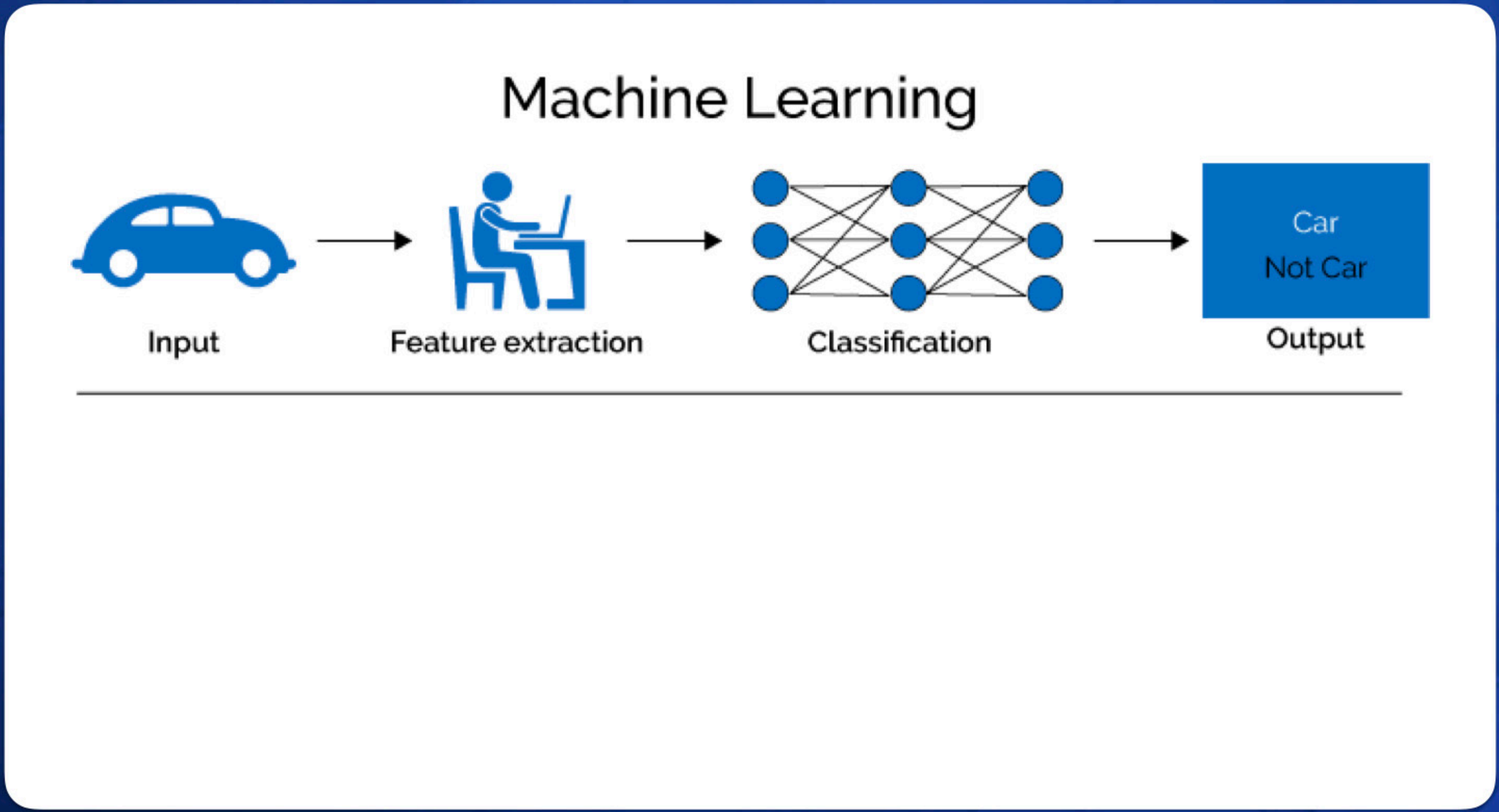


GO

- Unlike Chess or Checkers Go’s huge 19 x 19 board and nearly unlimited moves make it impossible to use pre-calculated AI strategies that look at every possible combination of moves.
- So the best computer programs written by humans only manage to reach amateur levels.
- AlphaGo built it’s own winning strategy by analyzing tens of millions of past Go matches and by playing against versions of itself millions of times to continue to hone it’s own decision making
- Go requires intuition and creative strategies to win, and thats what AlphaGo developed through a combination of algorithms supported by Deep Learning methods.

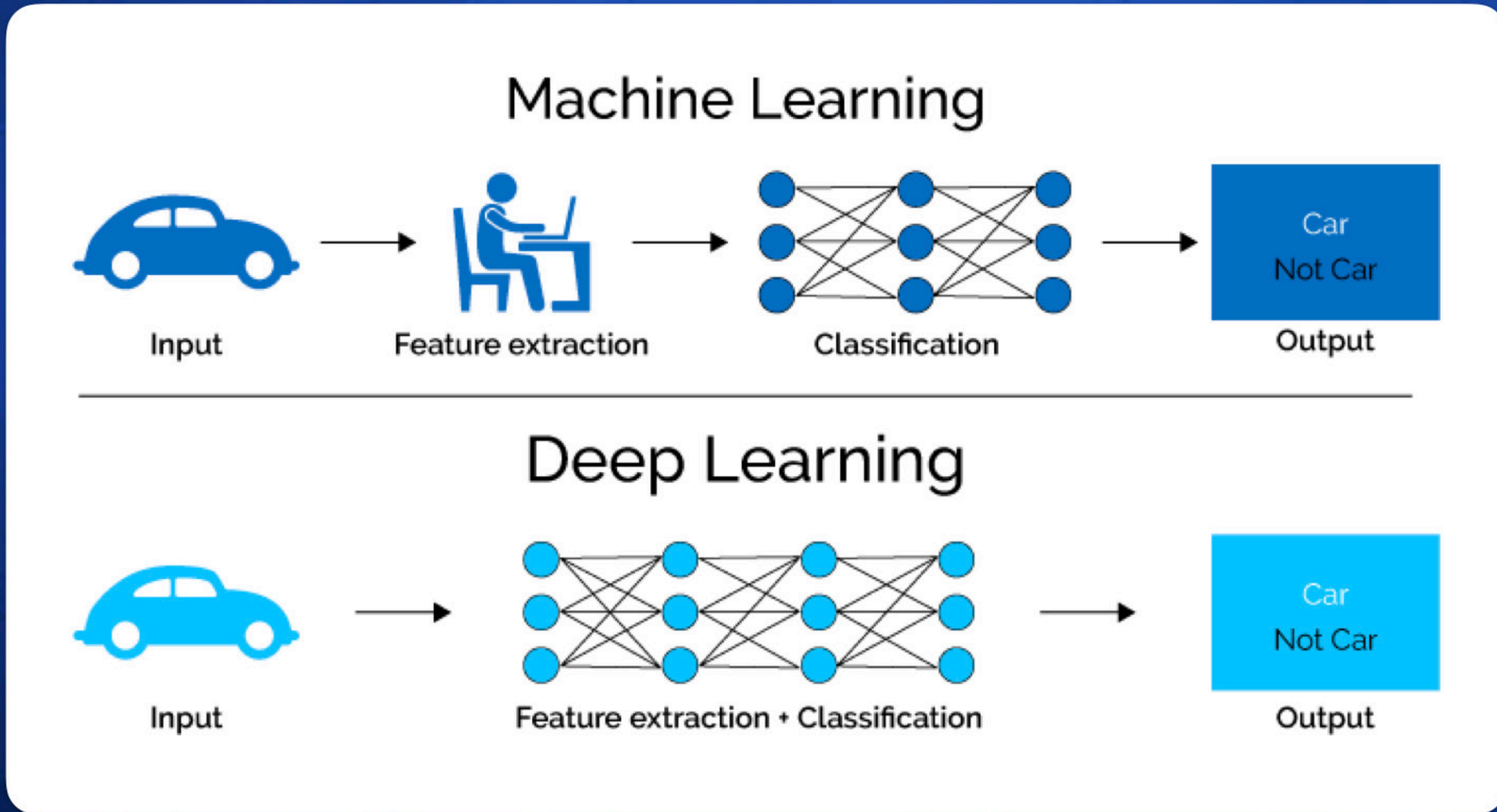


- So how does deep learning work?
- The best way to understand it is to contrast it with traditional machine learning.
- Let's look at this comparison in the next slide



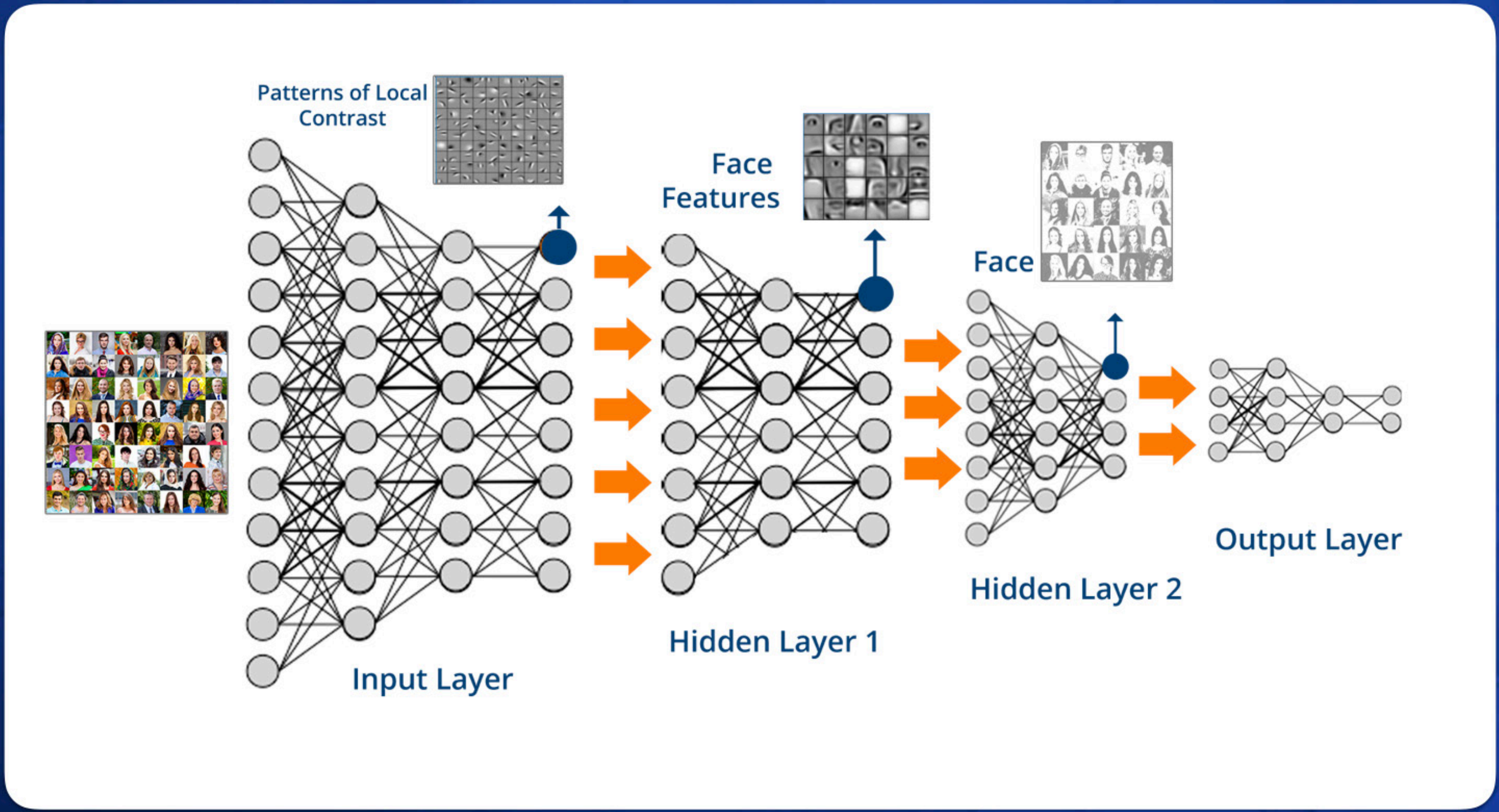
DEEP LEARNING VS MACHINE LEARNING

- In Traditional Machine Learning a human evaluates source data and performs a "feature extraction" step.
- Thats a fancy way of saying that she reviews the collection of historical data and determines the most meaningful fields or predictors to use.
- Our human then finds the most suitable Machine Learning Algorithm to use with those predictors and tests her hypothesis usually by trial and error, until a good combination of data predictors and algorithm(s) are selected to create the most accurate predictions she can find.
- In short, traditional Machine Learning uses "hand tuned" features processed through hand picked algorithms.



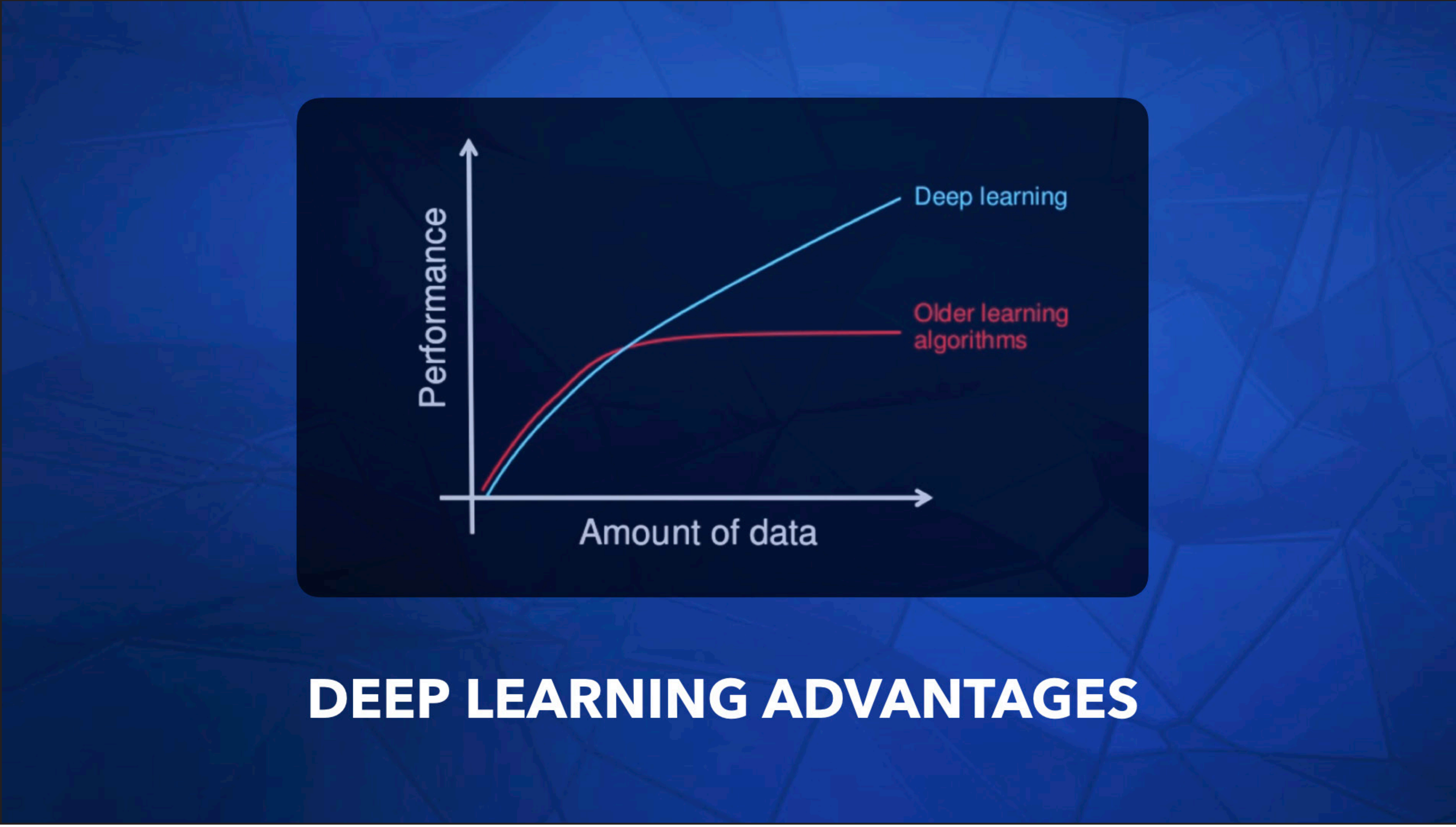
DEEP LEARNING ADVANTAGES

- But like Go, not all problems are so cut and dry.
- Sometimes we don't know whats important in our data, and sometimes complex problems don't fit in our traditional task-specific algorithms or sometimes they require multiple combinations of algorithms too long and complex to hand-select.
- Thats where Deep Learning excels. It processes raw Big Data sets to discover its own features that best represent the problem, while iterating on the solution at the same time.
- Our human doesn't even have to format the data let alone define data features.

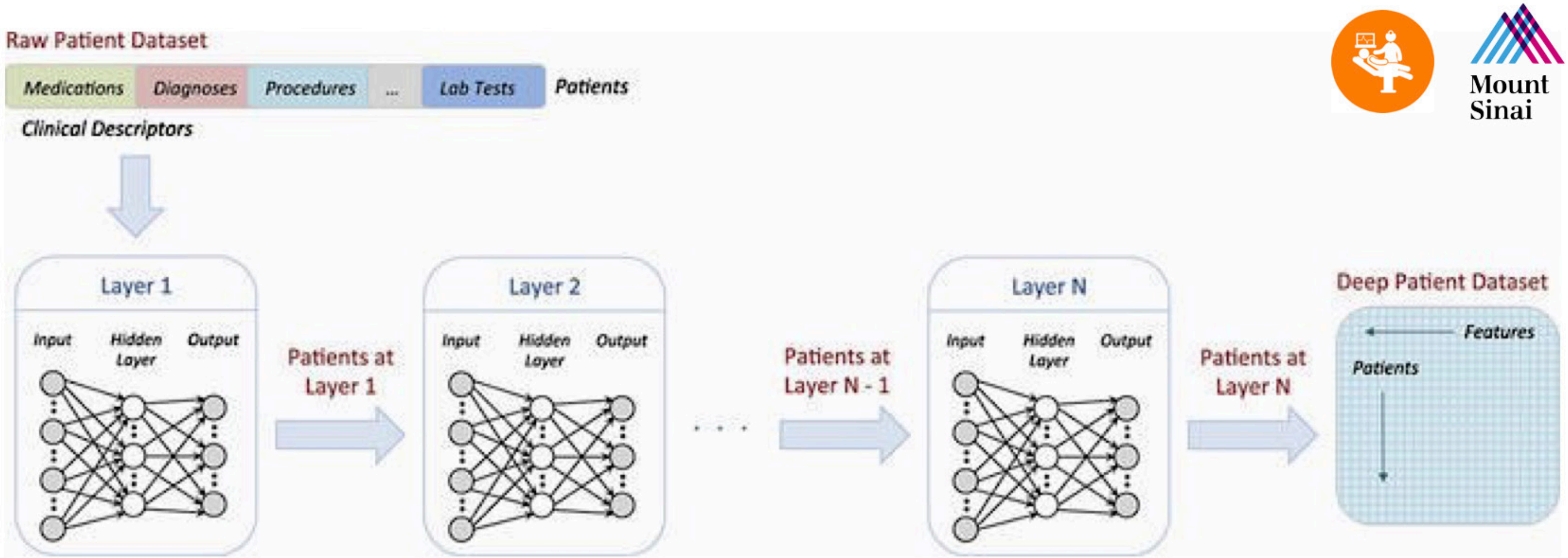


DEEP LEARNING ADVANTAGES

- There are many types of neural networks that behave differently but in general
- Deep Learning does this strictly by using concepts in data representation itself, to find the patterns in complex data.
- It processes the data through cascading layers of processing units called nodes,
- transforming the data at each node and identifying relevant feature at each layer until the full solution emerges.

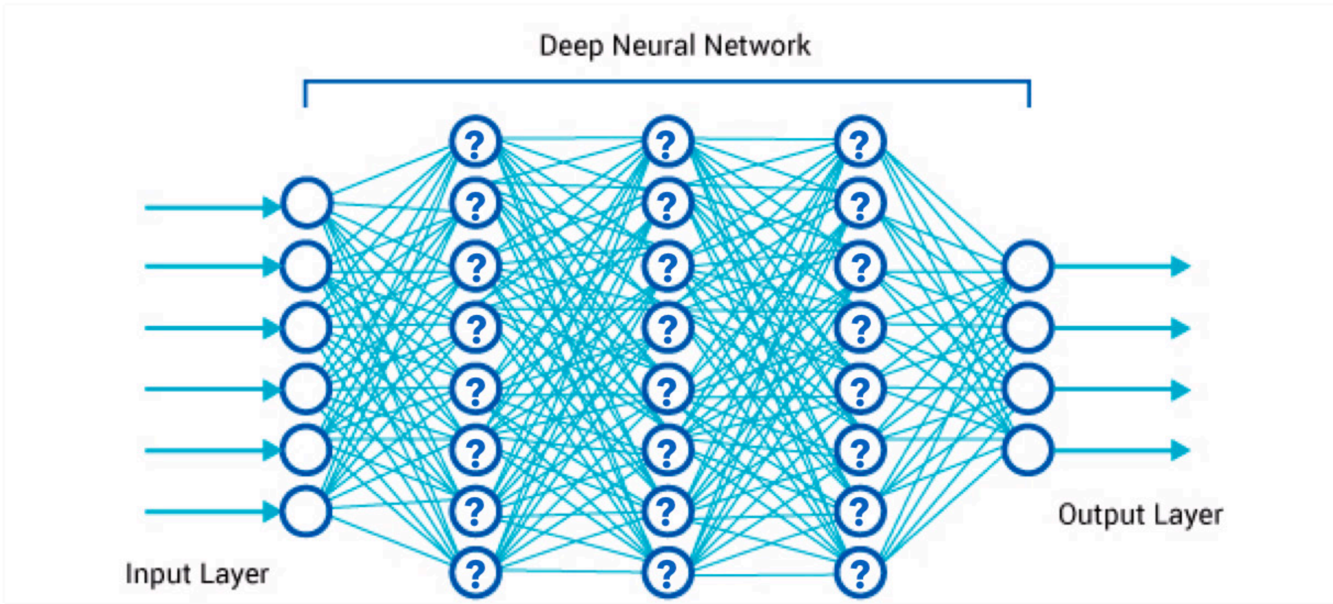


- One of Deep Learning’s biggest advantages is that the more data you throw at it, the better it performs.
- They’re designed for and excel with massive data sets.



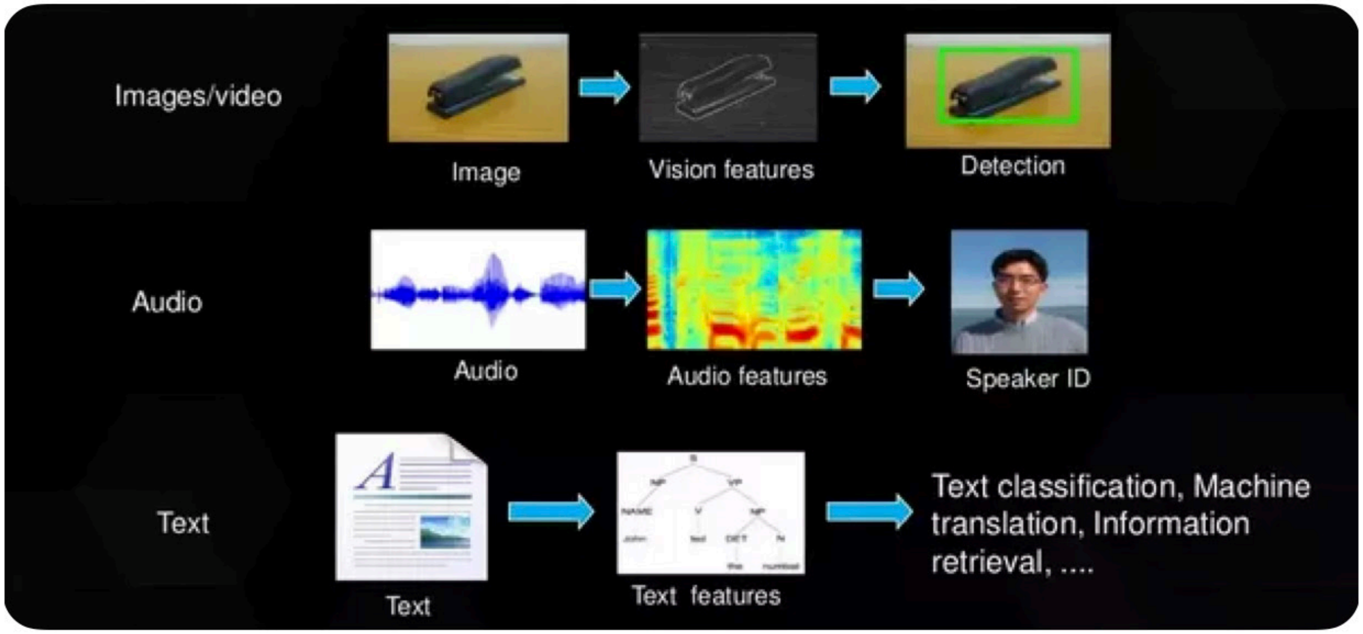
POWERFUL BIG DATA PREDICTIONS

- And when it comes to predictive analytics with Big Data the incredible predictive power of Deep Learning's Neural Networks can't be overstated.
- For example in 2015 a research group at Mount Sinai Hospital in New York, trained a deep learning network on the hospital's vast database of 700,000 patient records with hundreds of variables per patient.
- The resulting program, dubbed Deep Patient, was then able to predict a wide range of ailments on new patient records including the onset of schizophrenia.
- Researchers still don't know how the AI reached it's conclusions, only that they were accurate.



MORAL IMPERATIVE OF KNOWING

- Which also leads to the biggest criticism of Deep Learning:
- It's self generated solution formulas are almost impossible to interpret since they involve countless layers of complex mathematical transformations, too numerous to follow.
- This leads to the philosophical question: can we trust incredibly powerful predictive models, to make important decisions, if we can't understand the rationale behind them?
- On the other end of the spectrum, a simple Machine Learning formula can also be inappropriate due to it's lack of nuance.
- Interpretability vs predictive power can be a major trade off in predictive Machine Learning with Deep Learning truly being a black box option.



AI'S MOST POWERFUL TOOL

- But as mentioned, predictive analytics is just one solution Deep Learning offers.
- Deep Learning can be applied to unstructured data like text, sound and images which allows researchers to contribute diverse breakthroughs to other fields of computing.
- From Speech and Vision, Natural Language Processing, to Machine Translation and Chatbots, Deep Learning's neural networks enables many practical applications of Machine Learning and by extension the overall field of AI.

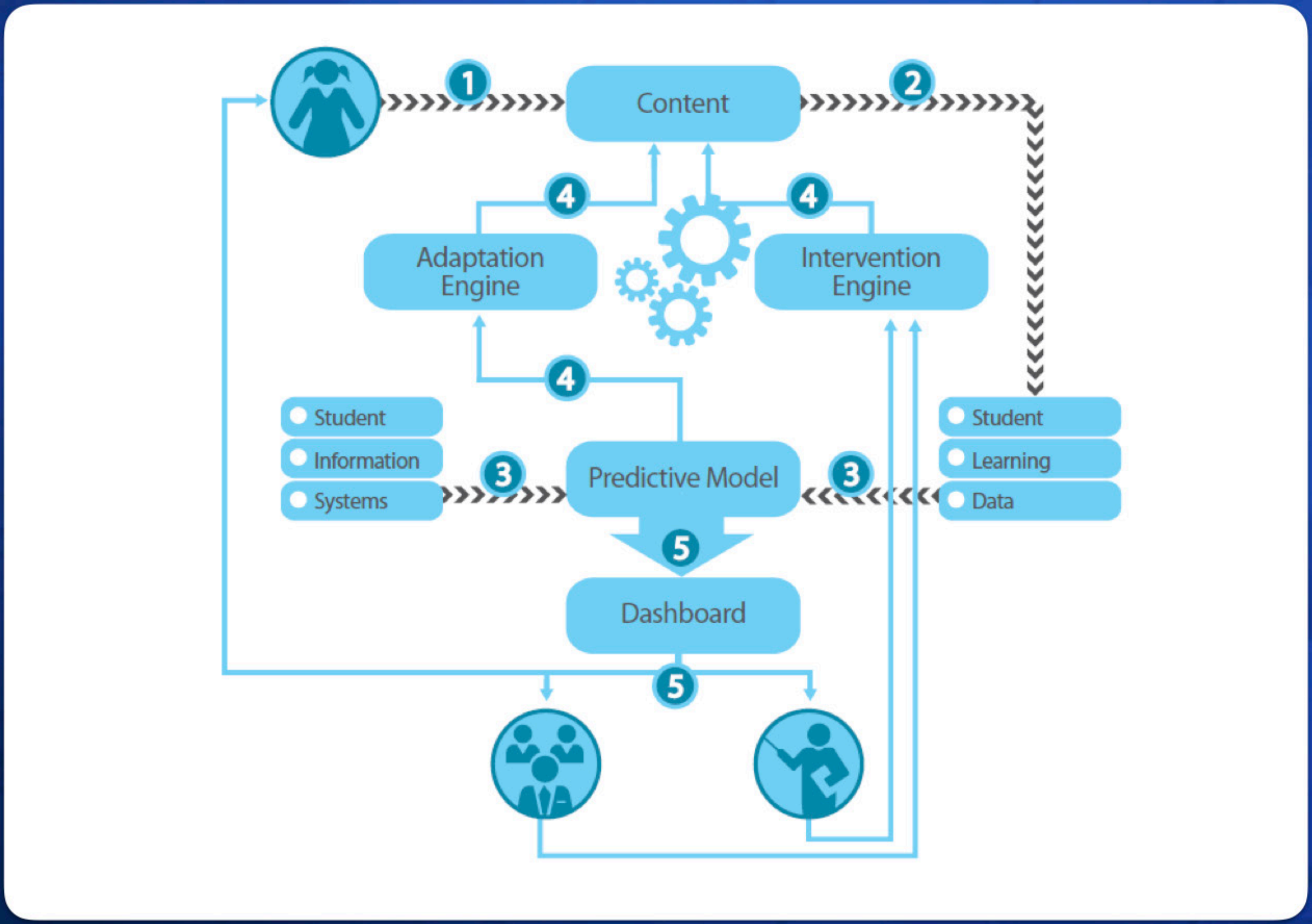
ROLE OF AI IN HIGHER EDUCATION

- Ok, interesting stuff!
- Now let's look at how AI and Machine Learning technology impacts Higher Education, specifically in our Post Secondary Solution areas.
- As Postsecondary experts you'll already be familiar with the solution areas themselves so we'll focus specifically on the AI-driven companies and technologies impacting them.

PS SOLUTION AREA 1:
DIGITAL LEARNING & COURSEWARE

- In our Digital Learning Tools & Courseware Solution Area we have Adaptive Learning Technology platforms powered by Learning Analytics.

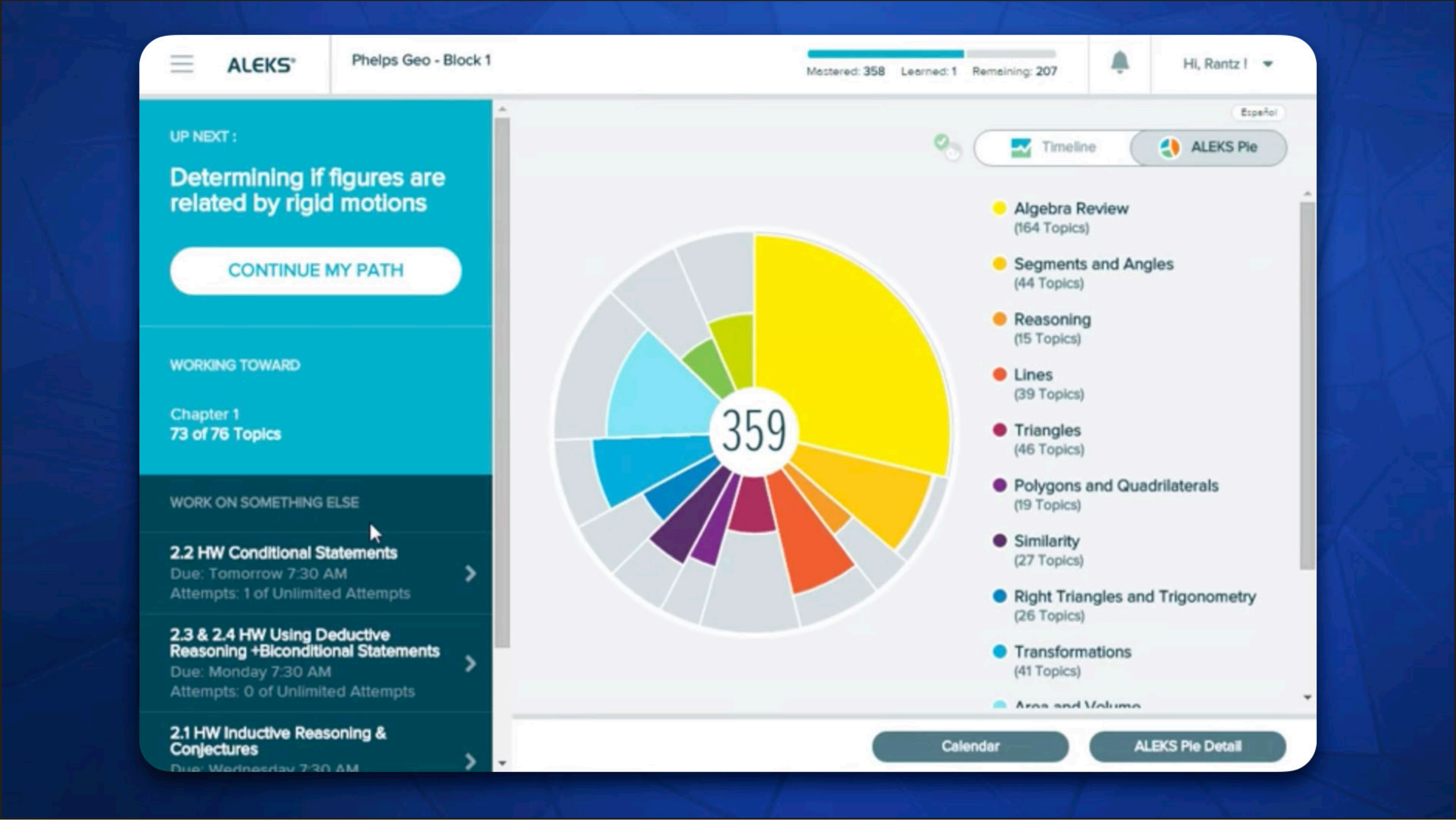
- Early and Entry-level Adaptive Learning Technology products tend to focus on determining the learner's pathway based on simple criteria such as the answer to a multiple choice question.
- They're often based on an underlying linear model that is too simplistic and not truly adaptive.



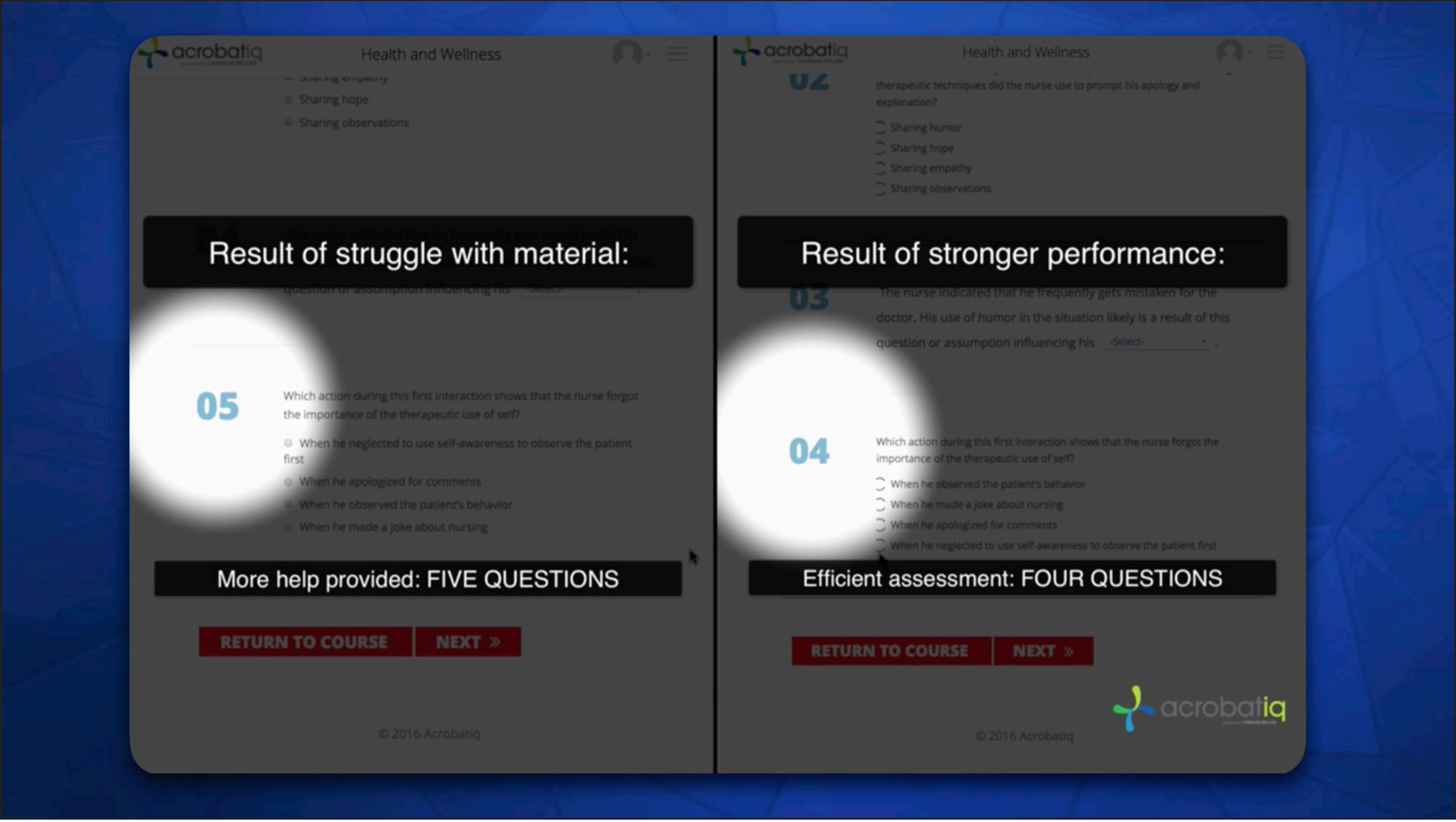
- More advanced tools utilize artificial intelligence methods for adaptability that look more like this.
- These systems are increasingly dependent on large-scale collection of learning analytics, outside of the actual learner.
- With more data we expect to see greater capitalization of Machine Learning models and potentially even Deep Learning which should make these products more effective.

- Here are a few noteworthy Adaptive Learning Platforms
- Companies such as Acrobatiq, CogBooks, Realizeit, and Smart Sparrow have all been working to prove adaptive learning's viability where instructors can build adaptive courses using their own content, thus making the course more of a personalized learning experience.

- I've highlighted companies that the Gates Foundation have already made grants to in yellow.
- And companies that have been acquired in pink.
- Aleks was acquired by McGraw-Hill in 2013 and Area9 was acquired by McGraw-Hill in 2014



- Of all these Digital Learning and Coureware platforms, Aleks is the oldest and longest running.
- It started in 1996 and is one of the first to utilizes big data to ascertain and implement efficient curriculum paths for individual students.



- Acrobatiq is among the newest platforms, first released in 2013. It's Smart Author platform generates real-time predictive learning estimates for each student against every outcome in the course.
- You can see it giving more questions to those struggling, providing more guidence.
- You should have an additional document detailing each companies location, investment rounds, and Content Models along with this presentation. Contact us for a copy if you don't already have it.

PS SOLUTION AREA 2:
TECHNOLOGY-ENABLED ADVISING

- Our second Post Secondary solution area is Technology-Enabled Advising.

PS SOLUTION AREA 2: TECHNOLOGY-ENABLED ADVISING

ellucian. EAB Starfish

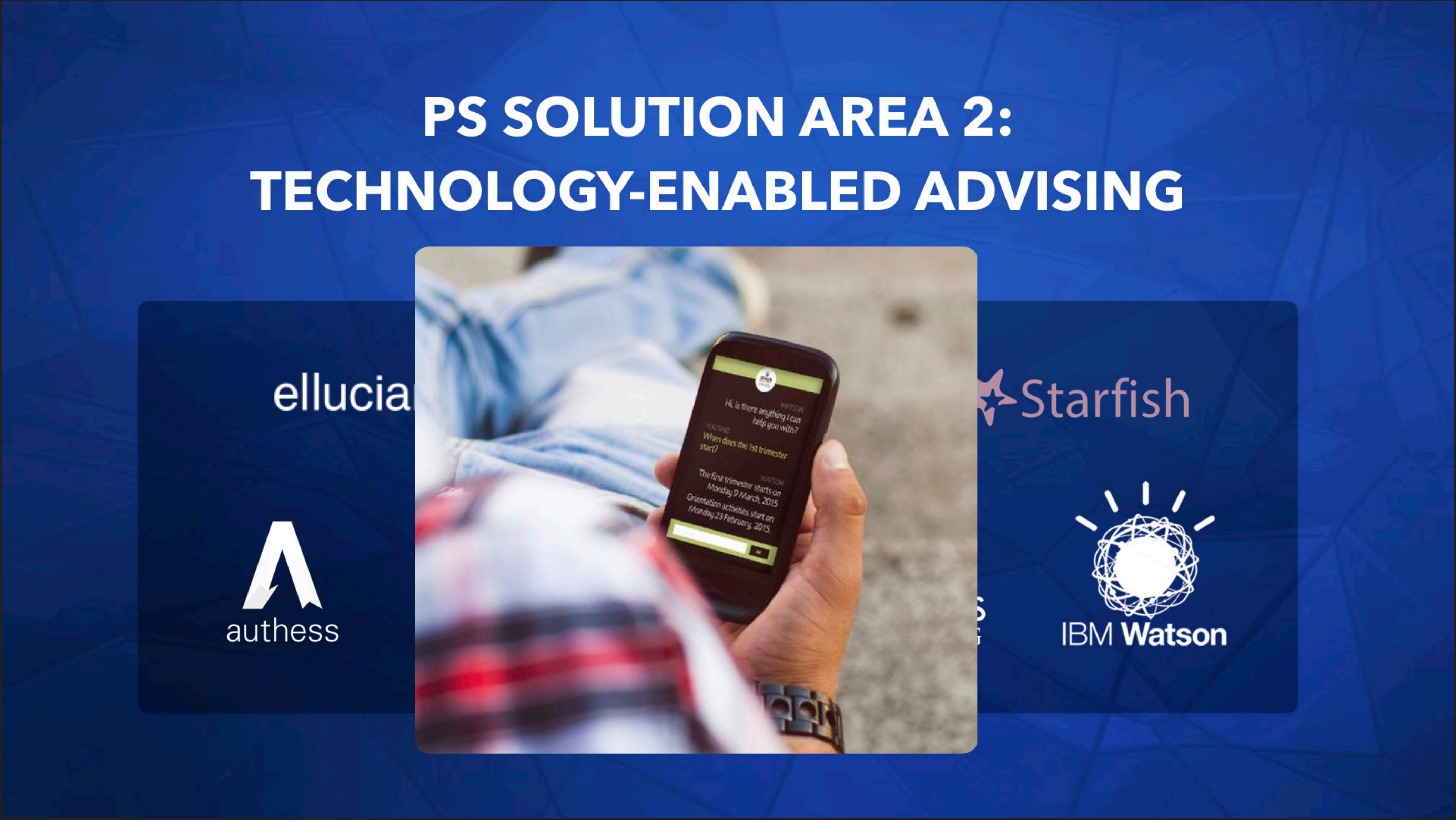
authess AdmitHub CIVITAS LEARNING IBM Watson

- This category in particular is incredibly well adapted for the solutions that AI's Machine Learning models can provide.

- Boston-based AdmitHub is worth a big mention. They advertise themselves as “Conversational AI for College success” and run the Georgia State University chatbot we mentioned earlier in this presentation.

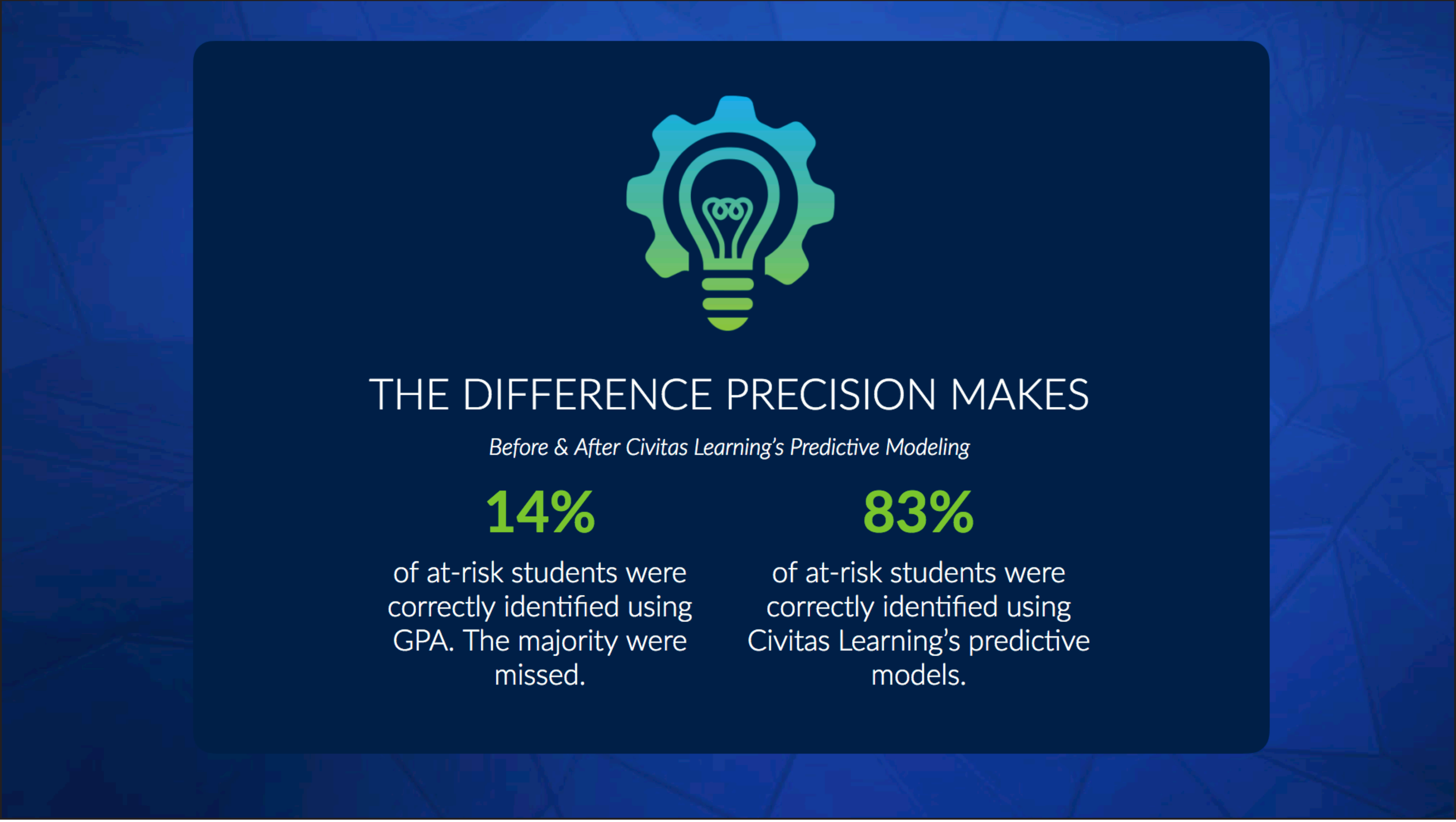
PS SOLUTION AREA 2: TECHNOLOGY-ENABLED ADVISING

- Not only do they support FAFSA assistance but also everything from orientation to immunization to transcript submission.

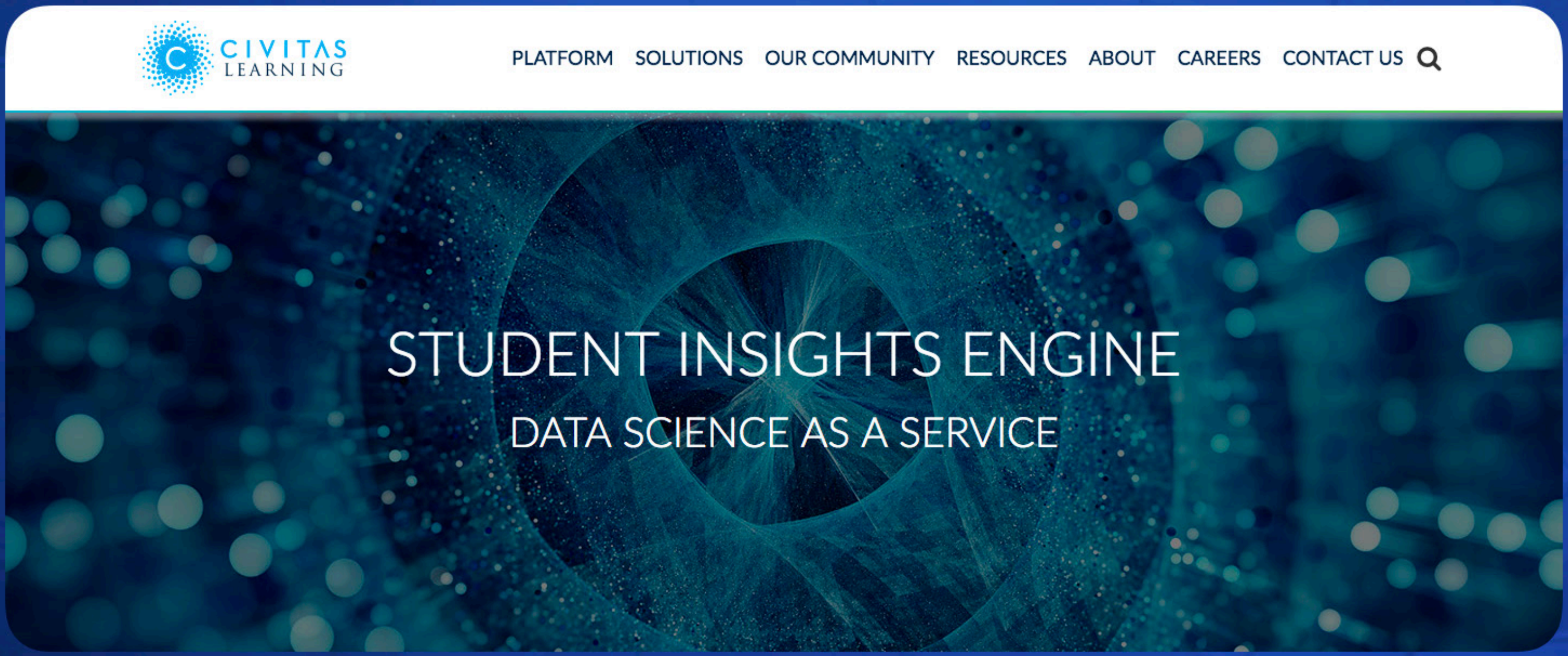


- Not to be outdone, Deakin University partnered with IBM to be the first worldwide to implement Watson Engagement Advisor, a Student Advisor Application version of IBM Watson.
- Watson is integrated with their platform DeakinSync with students reportedly asking up to 1600 questions a week to learn the ins and outs of life on campus.
- AdmitHub and IBM Watston are great examples of the indirect benefit of Deep Learning which has reducing word errors in speech recognition by 30%.

- A core staple of the Foundation's work in this area is in promoting IPASS or Integrated Planning and Advising for Student Success platforms.
- As you know, these platforms utilize the power of predictive data analytics, against institutional data sets to promote student success in 4 key areas:
- Progress Tracking, Student Planning (streamlined academic pathways), Advising & Counseling (Technology-enabled advising), and Early Alert Systems
- Many IPASS platforms like Civitas market on the sophistication of their predictive modeling

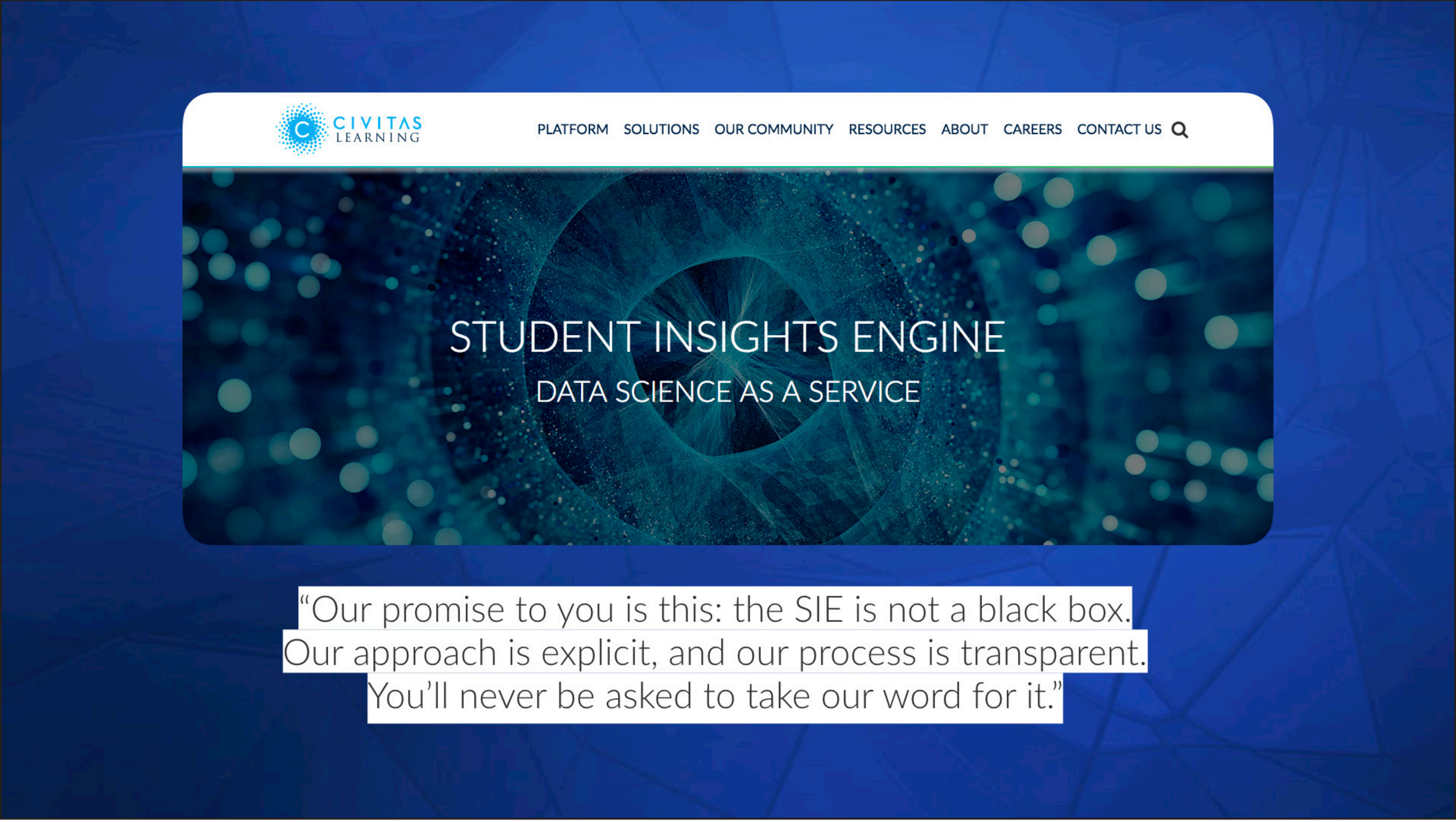


- In most cases these are Machine Learning Models that leverage known algorithms with their own custom classification techniques.
- These models are garnering some impressive outcomes.
- For example Civitas states 14% of at-risk students are correctly identified using traditional GPA-only risk detection where 83% of at-risk students were correctly identified using their models.

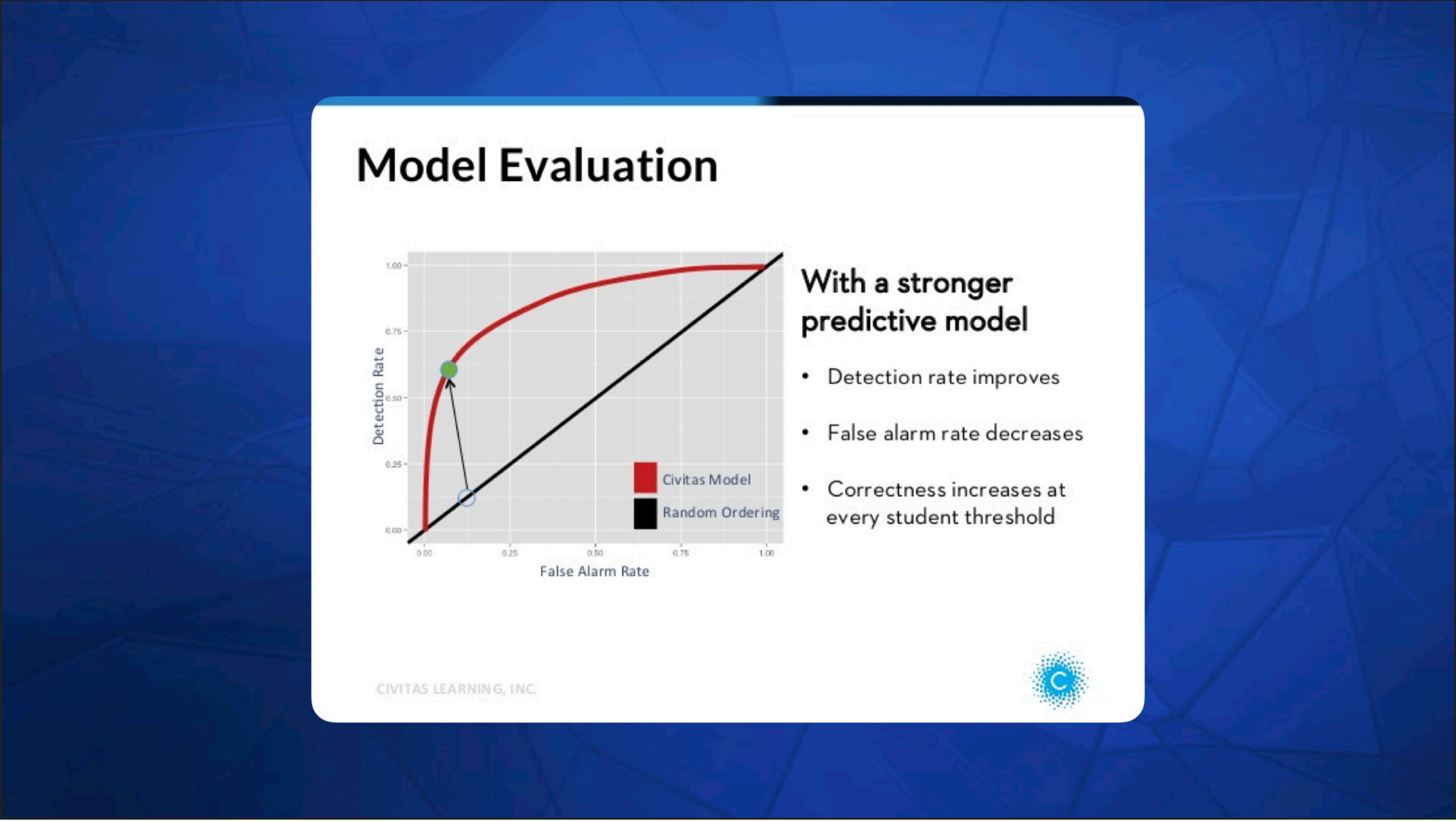


DEEP LEARNING FOR IPASS?

- One question is will IPASS products leverage the next level predictive power of Deep Learning?
- These platforms are already able to leverage an ever expanding anonymized dataset of analytics every time a new institutions signs up, so it would make sense.



- However many Machine Learning analytics based companies are already battling the stigma of black box predictions so if they are using them they don't seem to advertise it.
- For example, Civitas's portal for their Student Insights Engine contains this quote explicitly stating their solutions are not a black box and you won't have to take their word for it.
- No matter how powerful the algorithm, data science must still show proof to the institutions for the "Moral imperative of knowing".



- Companies like Civitas often do this by giving upfront visibility to their list of predictors, those data features we mentioned earlier.
- This is pretty critical as we now know that academic stats are not the whole equation.
- Civitas for example has noted that 99% of institutions in their sample were losing more students above 2.0 GPA than below it.
- The best predictive models today often include psycho-social contributing factors.

PS SOLUTION AREA 3: FINANCE / FINANCIAL AID

- Now let's discuss PS solution area 3: Finance & Financial Aid

PS SOLUTION AREA 3:
FINANCE / FINANCIAL AID

AdmitHub

NGV

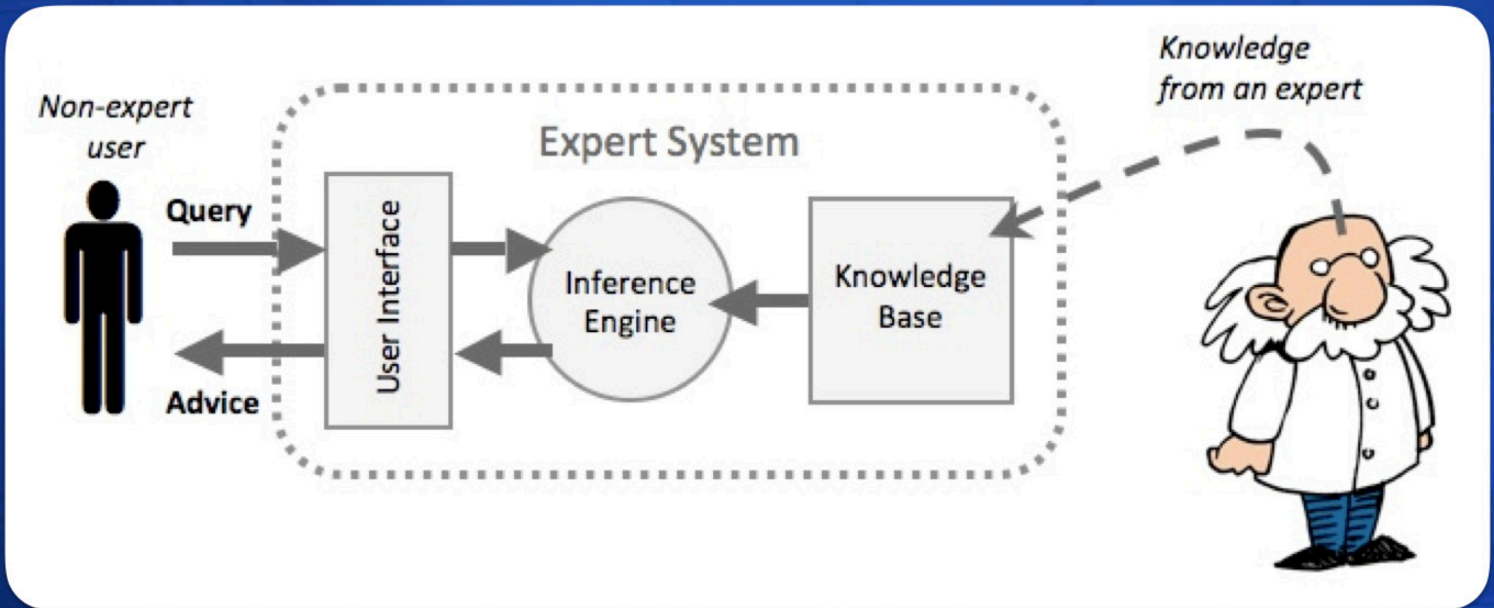
- Both AdmitHub and NextGenVest are text-based chatbot startups for helping students with financial aid.
- As previously discussed, AdmitHub is way more robust, supporting a lot more subjects other than finance while NextGenVest sticks with student loans and financial aid.

- NextGenVest has a concept called Money Mentors, that send deadline reminders, answer financial aid questions, curate scholarships, and more.
- Though they also use trained college students NextGenVest says they leverage machine learning and deep learning to “scale personalized communication for Gen Z.”

PS SOLUTION AREA 4: DEVELOPMENTAL EDUCATION

- Lastly, let's look at some ways AI and Machine Learning are leveraged in Developmental Education products.

- Carnegie Learning's Mika is an intelligent tutor systems for dev ed math.
- Mika teaches remedial courses for incoming college students giving a 1-to-1 tutoring experience.
- Mika delivers Hints and Just-In-Time Feedback to each student based not only on what a student knows, but also on the specific strategy a student uses to solve a problem.



AI EXPERT SYSTEMS FOR TUTORING

- One noteworthy field of AI that's making its way into Education is Expert Systems
- Also referred to Expert Context Advisors, Expert Systems are computer systems that emulate the decision-making ability of a human expert.
- They rely heavily on the AI field of Natural Language Processing for concepts like Automatic speech recognition (ASR), Sentiment Analysis, Entity Extraction, Natural Language Generation, and of course Question Answering.

- Our previously mentioned IBM Watson is the best known expert system.
- It's a question answering computer system capable of answering questions posed in natural language.
- The most practical impact has been in the medical field where IBM's "Watson for Oncology" application helps diagnose cancer and suggest treatments together with physicians.
- In research, the IBM "Watson Discovery Advisor" has aided a research team to analyze more than 70,000 articles for proteins links to cancer.
- In healthcare, 90% of nurses in the field who use Watson now follow its guidance.
- In education, IBM is currently working with both Pearson and Blackboard on leveraging Watson for one-on-one AI tutoring in their products.

- while on the topic of AI tutors another relevant field of AI is Affective Computing.
- Affective Computing Technologies sense the emotional state of a user (via sensors, microphone, cameras and software logic) and respond in kind, for example by recommending an activity that better fits the mood of the learner.
- The idea is to create context-aware, emotionally responsive machines that cater to even the most subtly communicated needs.
- Combined with online learning AI systems a computerized tutor can react to facial indications of boredom from a student in an effort to motivate or boost confidence.

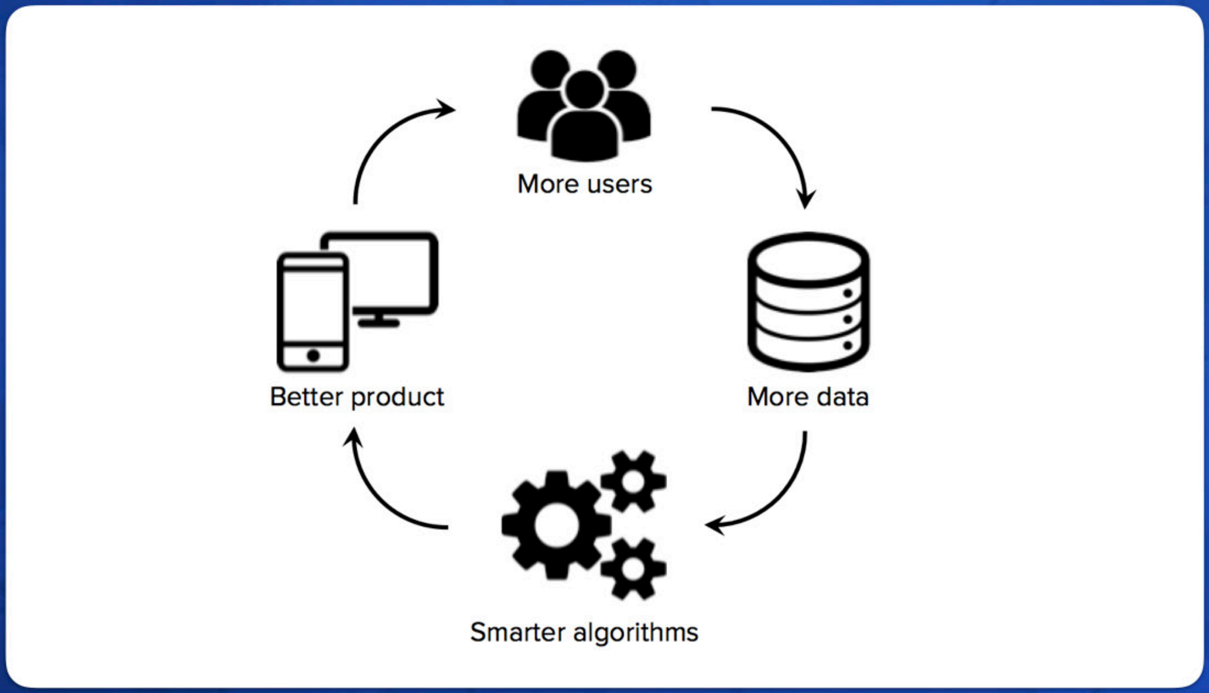
- In the same veign, Lenovo's AirClass is a new, distance-education and training platform, that uses webcam-based emotion detection and emotion analytics as one of it's core features.
- Affectiva Affdex is another company focused on real-time, learner emotional responses to educational content.
- Affectiva markets its technology as "emotion AI", using deep learning capabilities on an emotion data repository of nearly 4 million faces from 75 different countries.
- They're currently looking for Education Partners to collaborate with on the next generation of emotionally-aware education solutions.

BRIDGING THE EDUCATION TO EMPLOYMENT GAP

- On the other end of education, Artificial intelligence can play a key role in better connecting Education Systems and labor markets.
- Improved pattern recognition can pinpoint skill sets and traits that would enable someone to be successful in a job, and surface insights that have previously not been used.
- This can help detect promising candidates with less conventional credentials and free us from using school reputation as a proxy to assess potential.
- AI can also be used by governments to forecast detailed job-market demand more accurately and steer educational institutions to adapt their curricula and approaches accordingly making sure students have the skills required to fill those jobs. Saudi Arabia is currently exploring using Machine Learning in this way.

AI RELATED CHALLENGES IN HIGHER ED

- Now that we've learned about how AI and machine learning is impacting Higher Ed, let's discuss some AI-related challenges.



THE NEED FOR HIGH QUALITY TRAINING DATA

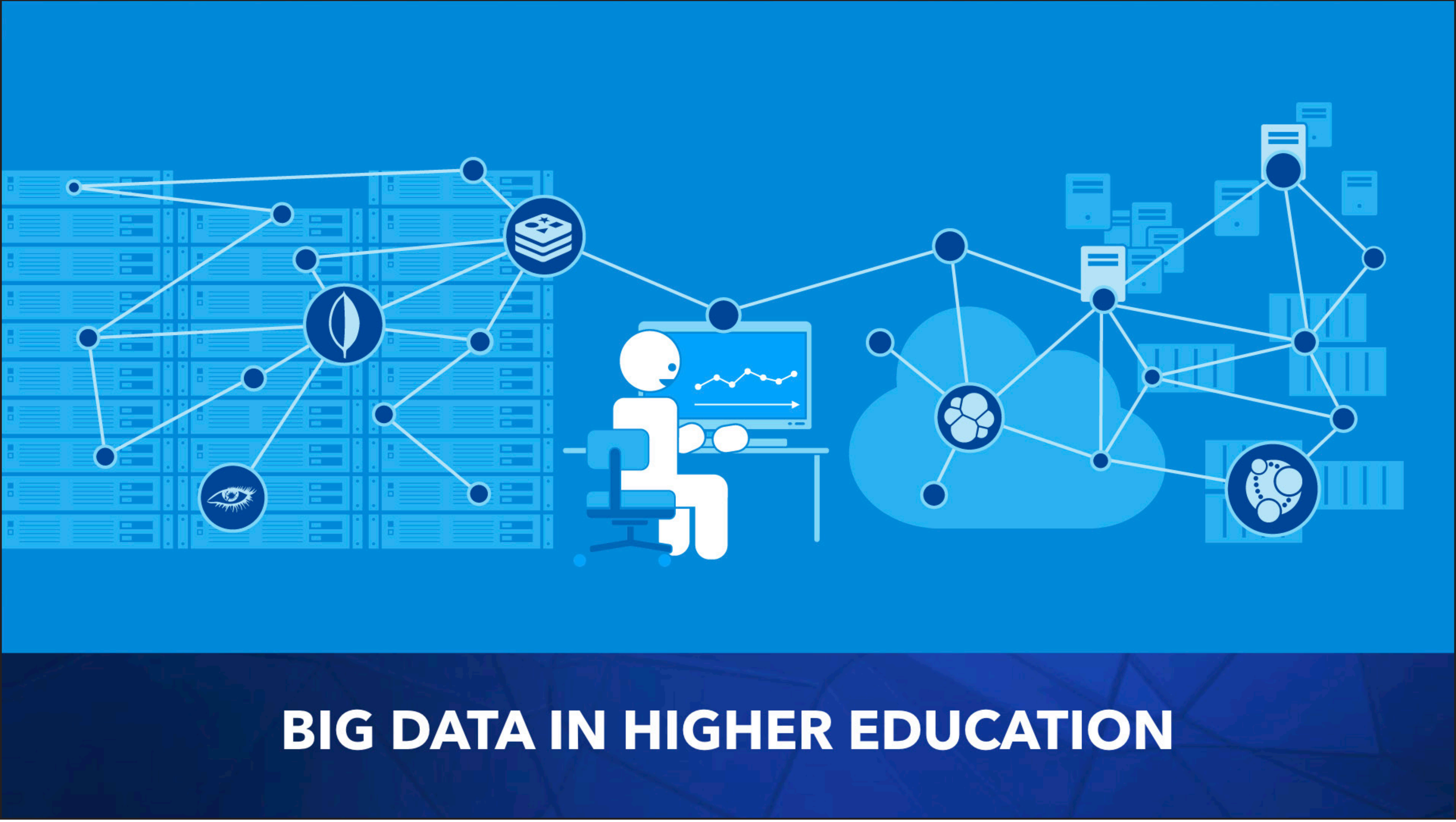
- Access to high-quality training data is critical for products that use machine learning as a core technology and thats no different for Higher Ed.
- Challenges for high-quality data not only include a lack of quantity, but data bias—that is data lacking good representation, or missing the right metrics.

(a) Three samples in criminal ID photo set S_c.

(b) Three samples in non-criminal ID photo set S_n

BIAS IN DATA AND MODEL TRAINING

- Like other forms of predictive analytics, Machine Learning algorithms are only as bias-free as the training data presented to it.
- For example in this controversial test from Shanghai Jiao Tong University in China researchers used a variety of machine-vision algorithms to study the photo IDs of around 2000 criminals and noncriminals.
- Researchers then found their neural network could then correctly identify criminals and non criminals with an unsettling accuracy of 89.5%.
- But look at these representative examples used. The data set is biased with no smiling white collar IDs in the criminals set. The algorithm weighed heavily on clothing which would include social-economic factors and subsequent biases inherent in prison populations in this absurd objective.



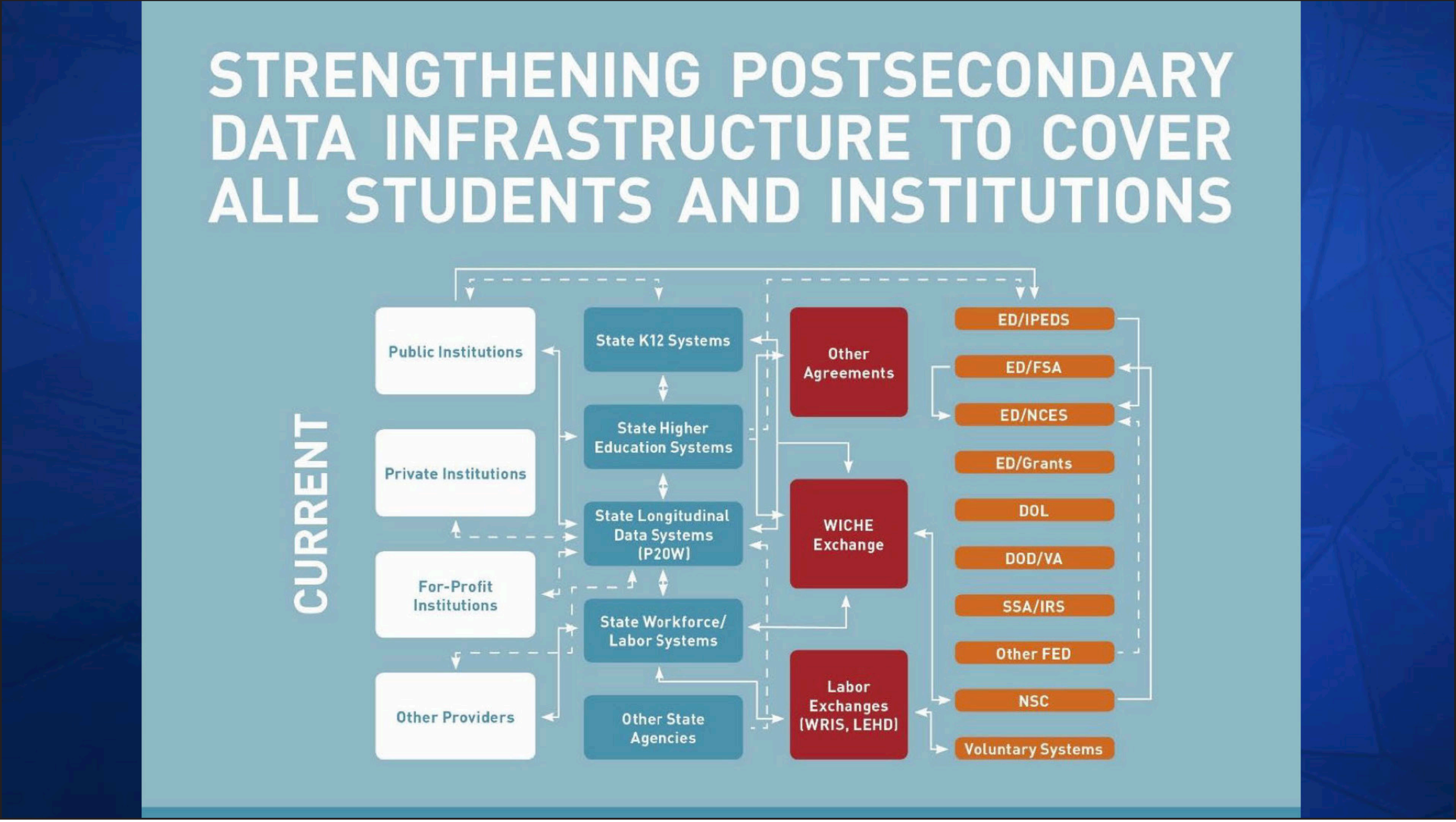
- keeping those biases in mind, in order for big data to be really effective, the dataset must be bigger than most institutions can collect on their own.
- Good datasets, however, are usually proprietary and hard to build and that's no different for Higher Ed.
- Since owning a large, domain-specific dataset can be a significant source of competitive advantage we can see a tendency for competing companies to keep data private.

PAR FRAMEWORK

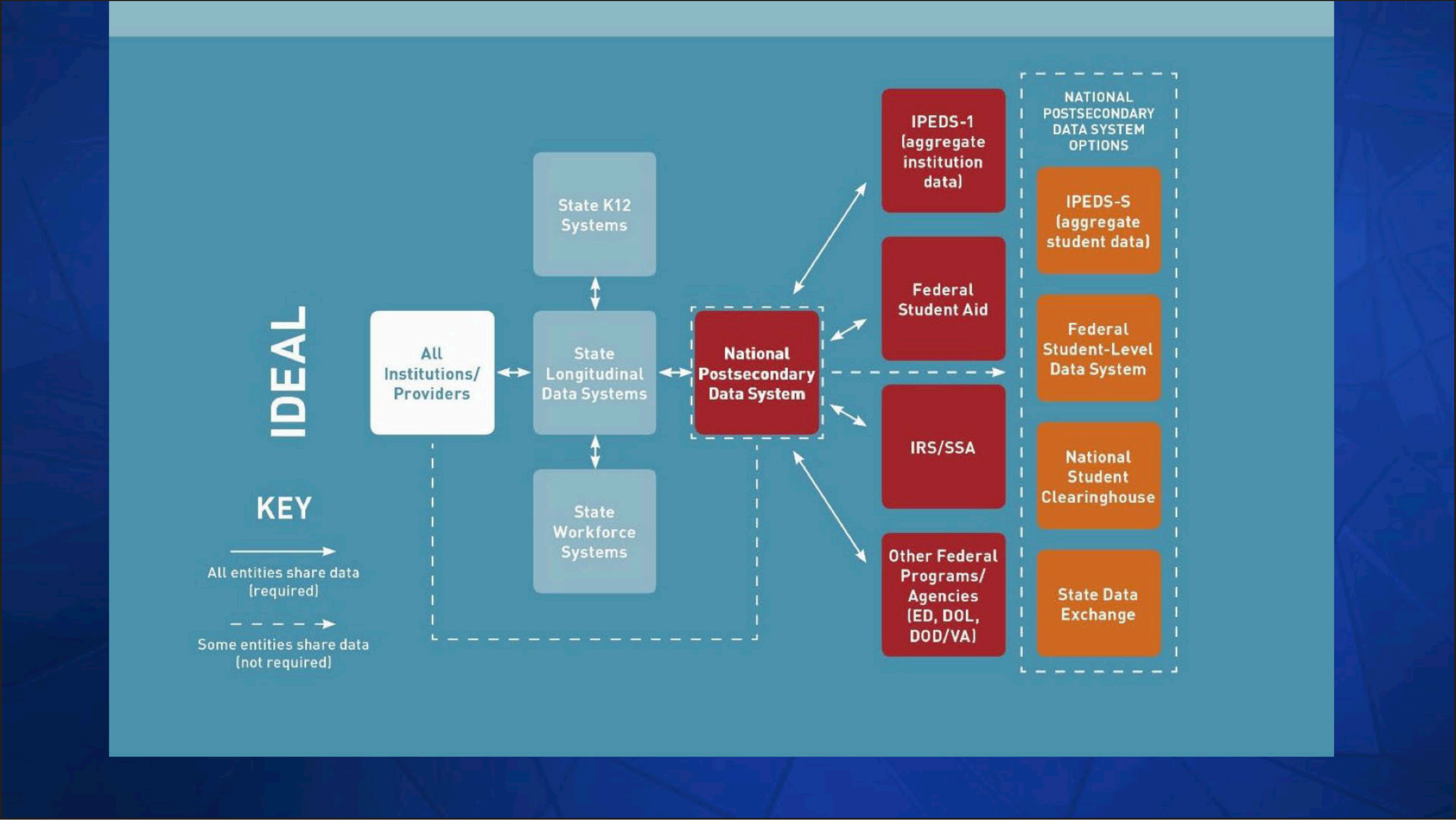
STUDENT SUCCESS ANALYTICS WORKING GROUPS

- An example of this could be seen in the non-profit PAR Framework.
- Many institutions have created data sharing working groups to promote the sharing of anonymized, Student Success analytics across member institutions
- and the PAR framework was one of the largest, combining data from hundreds of campuses and containing millions of student records, along with a predictive analytics framework to support it.

- Hobsons acquired Starfish Retention Solutions in 2015 and then the PAR framework in 2016,
- They combined the two into the Starfish Enterprise Success Platform.
- The previously non-profit PAR had spun off from the Western Interstate Commission for Higher Education, a Gates Foundation funded non-profit.
- It now remains to be seen how Hobsons manages this heritage and if it continues to be a competitive advantage over commercial competitors such as Civitas Learning who independently grew their own dataset.



- An ideal place to look for the Big Data required to power good predictive models is a centralized Higher Ed infrastructure.
- But this is how this infrastructure exists today.
- It features inconsistent, incomplete, and often conflicting metrics silo'd off between a series of disconnected systems.
- These systems were created at different times for different purposes.
- And the metrics published today often only include "traditional" student metrics in their datasets which ignore the new normal in higher education, New Majority students.



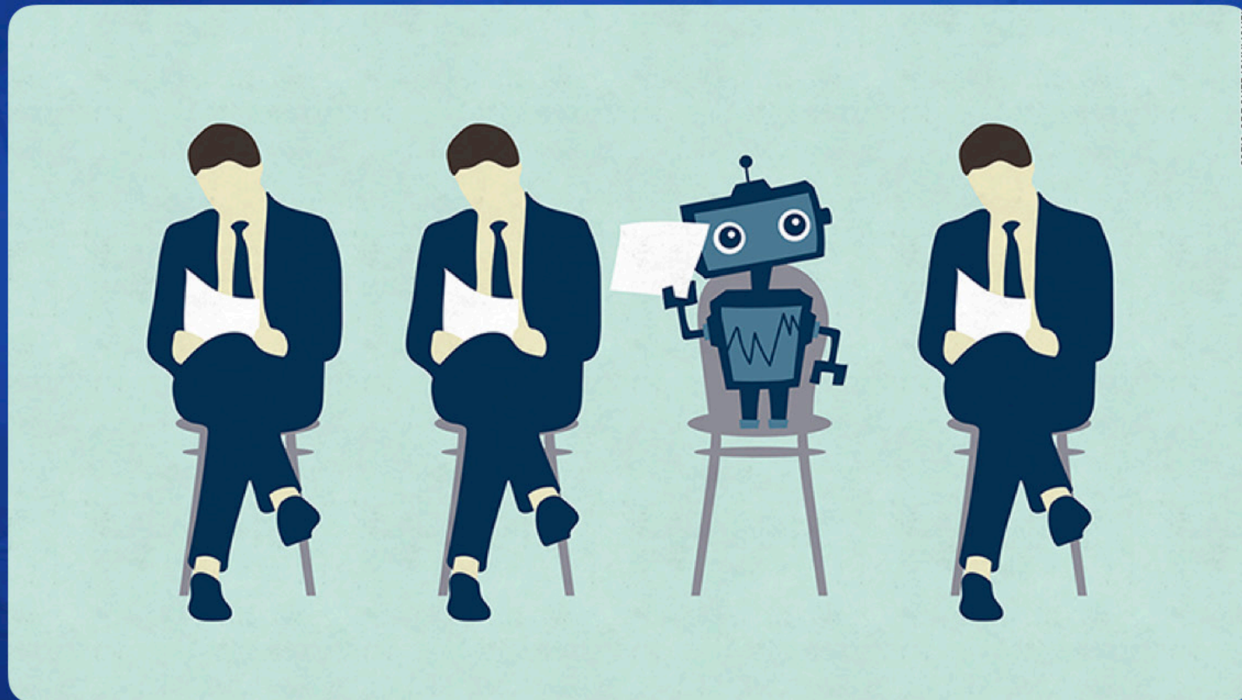
- As published in the Foundation's "Postsecondary Success Advocacy Priorities" -
- This is what the recommended "National Postsecondary Data System" would look like.
- A central collection of modern key performance metrics for all students in all institutions.
- If modern Data Scientists could get access to data like this, they could find patterns and answers to questions we haven't even asked yet.

- The foundation has partnered with "the Institute for Higher Education Policy" (IHEP) to develop a version of this framework with consensus metrics from leading institutions and other Gates-sponsored organizations.
- These metrics have a special focus on underserved populations who constitute two-thirds of higher education enrollments but are missing or obscured in current higher ed performance measures and public databases.

CHAMPIONS OF ACCESS AND SUCCESS
20 YRS
IHEP
INSTITUTE FOR HIGHER EDUCATION POLICY

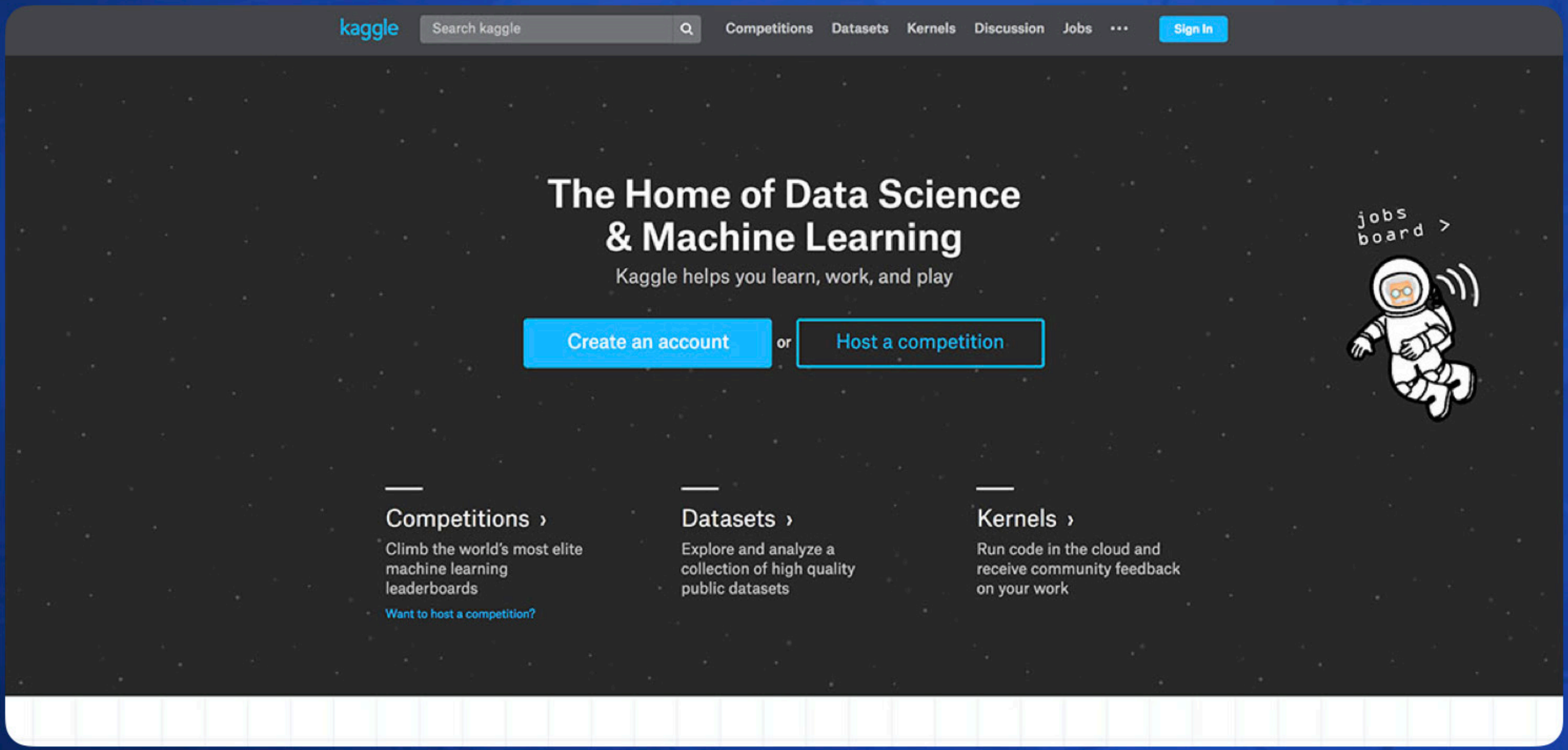
POSTSEC
DATA

- This new national postsecondary data system has been difficult to put in place as data sharing, privacy, and security are not yet governed by a coherent set of policies.
- They also present ethical issues, starting with who owns data on students, who can see it, who can use it, and for what purposes.
- Meanwhile IHEP and it's "PostSec" Data Collective of more then 40 partner organizations continues to publish policy papers and promote evolving student-level data legislation including an interim step of enhancing the utility of existing data systems
- This includes the Integrated Postsecondary Education Data System (IPEDS) and the National Student Loan Data System (NSLDS).



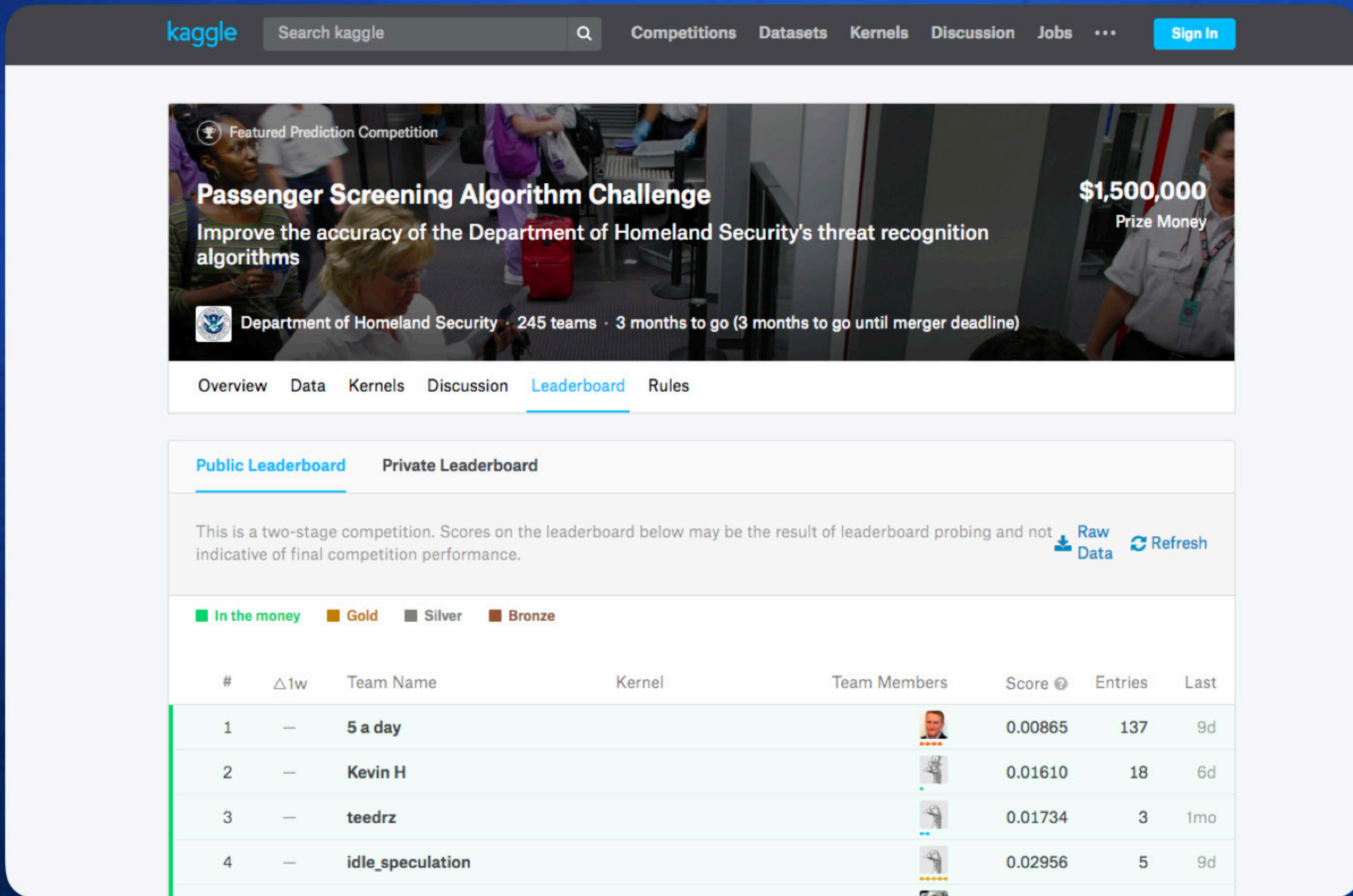
CHANGING JOB LANDSCAPE DUE TO AUTOMATION

- The growth of Artificial Intelligence promises benefits, but also poses some urgent challenges to jobs as well.
- The World Economic Forum recently projected that automation will eliminate at least 5 million jobs, worldwide, by 2020.
- Factoring AI-automation in skills training is becoming more critical as AI-technology becomes more u-biquitous.
- The workforce needs to be reskilled to exploit AI rather than compete with it. For students, a future-focused curriculum is a necessity.
- The World Economic Forum identifies 16 skills that are needed in the 21st century—including creativity, collaboration, initiative, and adaptability—but those are not included in standard curricula.

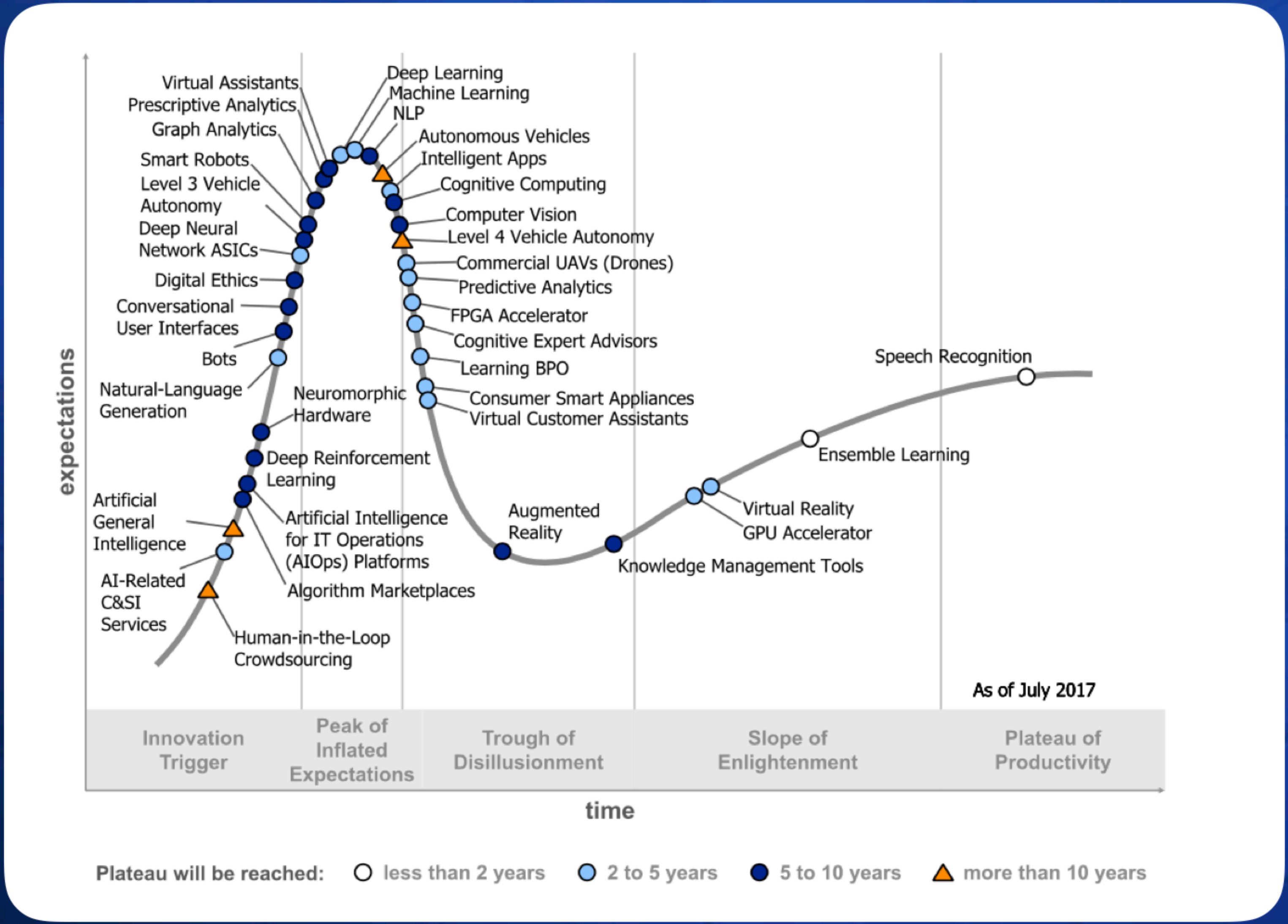


LEVERAGING THE DATA SCIENCE COMMUNITY

- Lastly let's talk about opportunities in leveraging the data science community.
- There is now a large, thriving Data Science community actively solving complex problems with Machine Learning and Deep Learning.
- Kaggle.com which was acquired by Google, is one of those communities that feature major Algorithm challenges sponsored by companies, organizations, and even the US Government.
- Of the 17 active competitions underway, the largest is a "Passenger Screening Algorithm Challenge" by the US Dep. of Homeland Security



- The TSA is challenging the broader data science community to help improve the accuracy of threat prediction algorithms using datasets of images collected from their scanners. 242 teams are currently competing for $1.5M in price money.
- We see this as a great potential avenue for the Foundation to tap into the broader Data Science community
- to potentially contribute solutions for our complex problems in Higher Education.
- With a single challenge grant contest we could have hundreds of research teams crunching our data to find effective Machine Learning predictors and solutions.



- We close with Gartner's 2017 "Hype Cycle for Artificial Intelligence" where you can see many sub branches of AI technology we discussed today such as Expert Advisors, Predictive Analytics and most significantly, Machine Learning are trending.
- Gartner estimates 2-5 years for these technologies to hit mainstream but as you can see, they are, in a large part already here,
- powering the advanced algorithms inside our phones and websites.
- Though AI like machine learning and deep learning are already starting to make an impact in education today, improvements in data standards, collection, and data privacy are probably the single biggest roadblock for AI to completely transform the Education space. It is sure to happen sooner than later.

THANK YOU

MAUVIS@POLYOLYGON.COM
MARK@POLYOLYGON.COM

Funded by:
BILL & MELINDA GATES foundation

- Thanks for watching our presentation.
- Questions? Comments? Reach out to us at mauvis@polyolygon.com and mark@polyolygon.com

SEPTEMBER 2017 ADDENDUM

- This research presentation was originally presented on September 14, 2017. We'll be periodically adding noteworthy developments in this space in subsequent months. Here are some noteworthy developments (up to end of Sept. 2017)

Azure
Machine Learning

Visual Studio

Excel

MICROSOFT LAUNCHES NEW ML TOOLS (EXCEL)

- Along with launching new Machine Learning tools and services on their Azure Cloud Computing Platform and Visual Studio for developers, Microsoft also announced a host of new AI functions right inside Excel at their Ignite Conference (Sept 2017).
- Excel will be able to support JavaScript and service APIs to dynamically pull in new columns recommended by analyzing and understanding your current data types. For example Excel could automatically detect a column is a list of city or company names and provide supporting columns like stock data or population data.
- Excel will also be able to pull in machine learning models directly to analyze data, generate insights on trends and outliers, and even provide data visualization through it's "Insights" tool. This update will arrive in early 2018 and is a strong indicator of Machine Learning's reach into the software we all use every day, not just for data scientists and developer tools.