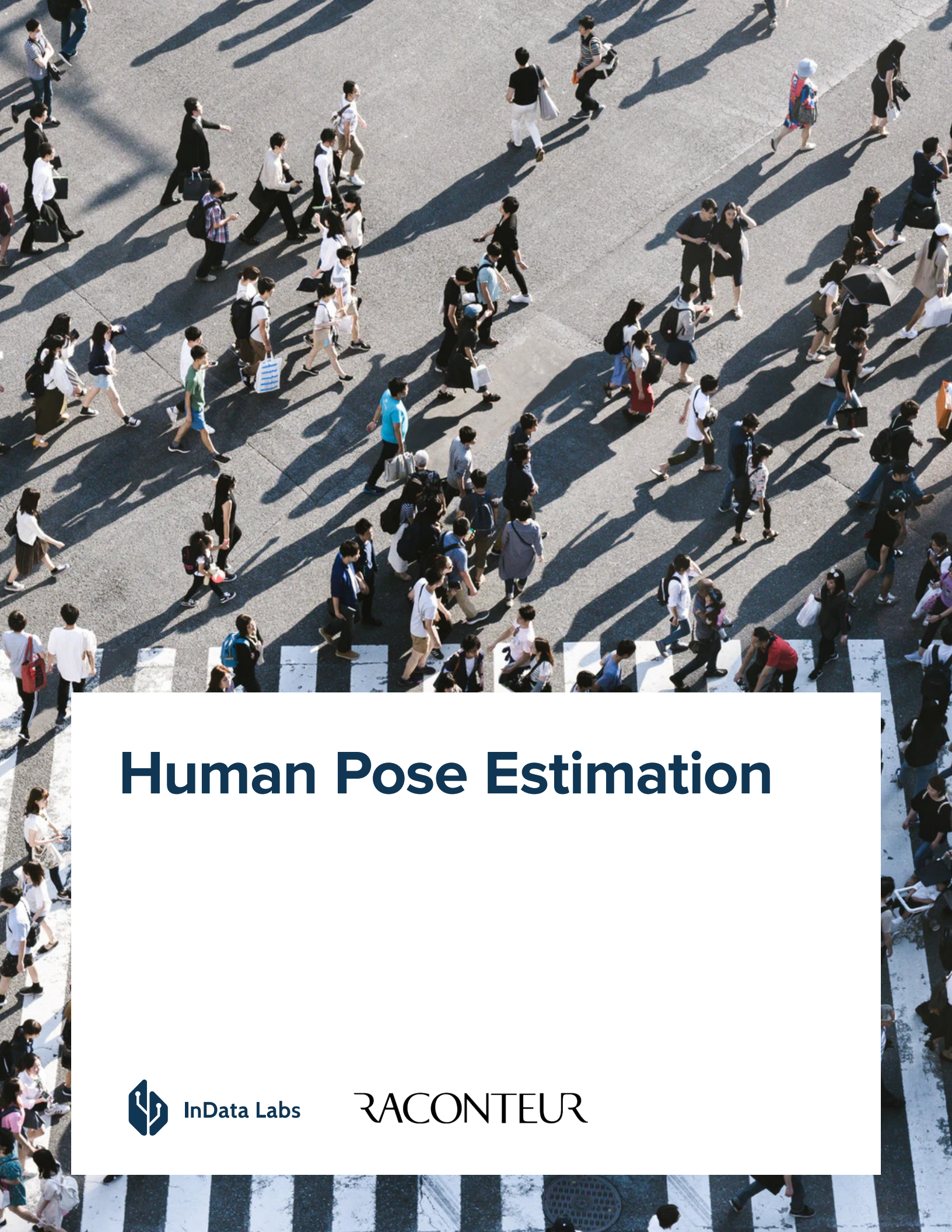


# Human Pose Estimation

InData Labs

RACONTEUR

# What is human pose estimation?

**One of the most mature artificial intelligence technologies, human pose estimation, has been quietly making its presence felt in business applications**

Currently hovering at the edge of mainstream adoption, many believe 2020 will see human pose estimation experience widespread growth in an array of new applications.

So why is human pose estimation so important? Put simply, it's easy for adults to tell the difference between different types of movements: we can effortlessly distinguish between someone breakdancing and doing push-ups, for example. Computers, however, see pictures as a number of pixels, so it's more difficult for them to understand what's happening in the picture or video.

Pose estimation is a machine-learning technique that's been developed to help a computer understand what position a human has taken, to analyse their movement further for a wide range of uses and applications.

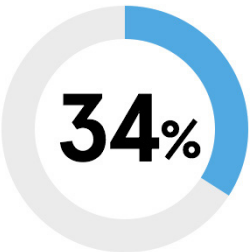
# How does it work?

First, let's look at how human pose estimation works. The easiest way to estimate the human body position in a picture is to determine key points on the person, such as knees, elbows, hips, feet, shoulders and neck.

"Pose estimation is usually run on 2D images or individual frames of a video sequence," Dr Rob Dupre, an expert in crowd counting and behaviour analytics, explains.

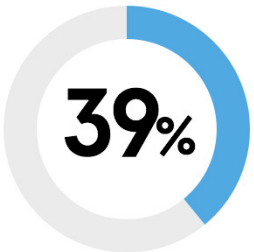
"The joint positions can be connected to form a rudimentary skeleton which can then be visualised on the image. The skeleton is usually presented in 2D on top of the image; however, another area of pose estimation research looks at projecting those joints into 3D, allowing for applications like motion capture for moving 3D CGI [computer-generated imagery] models."

Dupre says this kind of key point detection is comparable to the technology which supports programmes, such as Snapchat filters or Apple's Animoji, although they focus on just key points associated with the face.




of global senior decision-makers with influence over IT say their companies have deployed AI

*IBM 2020*




are ramping up exploratory phases with AI

Now that we can detect and identify where a person is in a scene, the next question is what are they doing?

Pose estimation, through its ability to detect joint positions, can help find the answer. Using its additional metadata, we can infer a host of new information from what a camera is seeing. The relationship between these points allows us to detect specific body positions.

"For example, if a person's ankles are horizontally aligned with their hips and their head, we might be able to infer that they are lying down," says Dupre. "Using tracking technology, we can check to see if that same person was lying down in a previous frame. This might indicate they have fallen down and we might want to alert someone to this."

# Why is the technology set to take off?

Detection of people has long been a major focus for a range of industries, including security, business intelligence, health and safety, and entertainment. With recent developments in machine-learning algorithms, the accuracy of these detections, and the hardware requirements to run them, pose estimation has now reached a point where it has become commercially viable.

In addition, the technology's growth is also being driven by the coronavirus crisis, says Dupre.

"Businesses have had to respond to stringent requirements on capacity and social distancing which has typically been addressed through manual monitoring and counting. The cost to do this at scale has generated an instant need for accurate automation of these tasks and pose estimation is one technology that provides this as well as much more," he says.

Indeed, social-distancing applications, combining human pose estimation and distance projection heuristics, now exist that encourage individuals to maintain physical distance from each other during the pandemic.

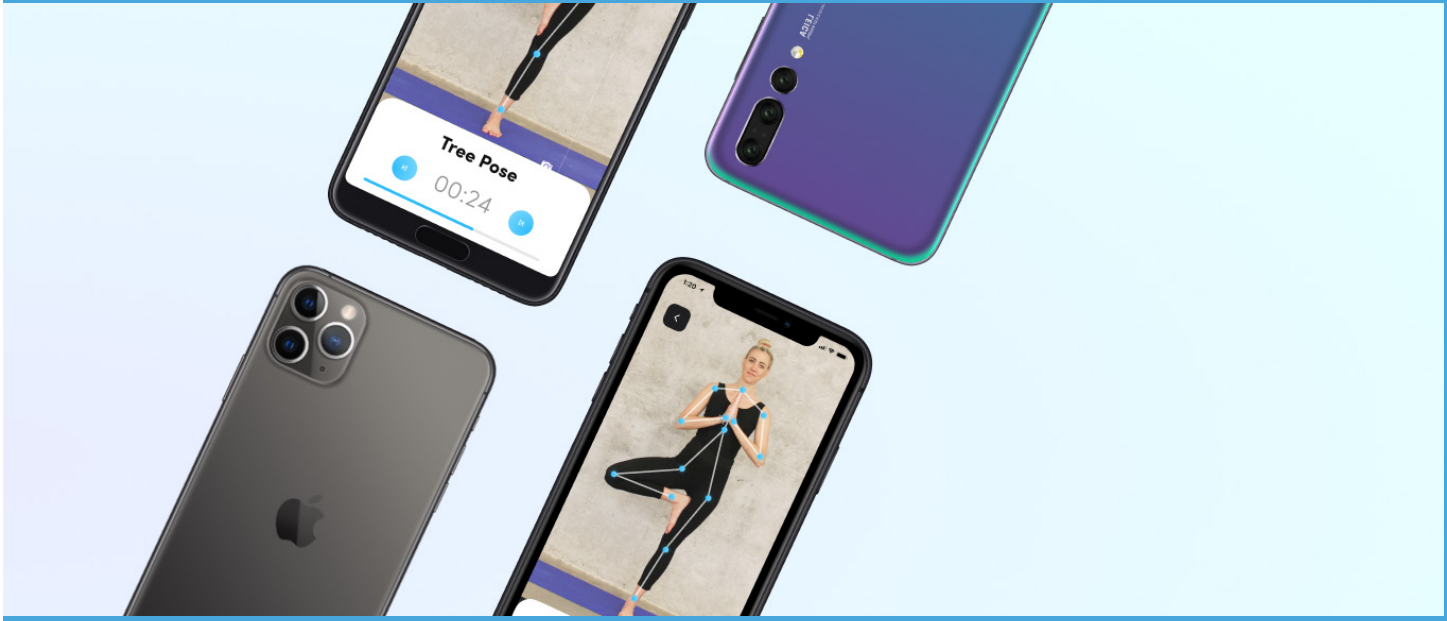
It's also worth stating that pose estimation is compliant with the General Data Protection Regulation as it doesn't identify personal features; it simply uses key points on the body to identify a human presence.

IDC describes the convergence of technology and humanity, as augmented humanity. It notes: "Technologies are generating new human-like experiences, leveraging a new range of human-machine interactions that are three-dimensional, more intuitive and more natural."

Human pose estimation will help enable a new generation of interaction and intersection between the physical and digital worlds.

Social-distancing applications, combining human pose estimation and distance projection heuristics, now exist that encourage individuals to maintain physical distance from each other during the pandemic

# Human pose estimation applications




Human pose estimation is lauded as especially important in a wide range of use-cases.

One of the more widely known examples of pose estimation was with Microsoft's Kinect hardware for the Xbox. This allowed players to interact with a game using their body without the need for the player to hold anything for the system to track, as now commonly seen in virtual-reality applications.

Kinect tracked 25 joints across the full body. However, this technology relied on specialist camera technology that limited the range. Modern pose estimation techniques only require a standard RGB (red, green and blue) camera, making the technology far more applicable.

"This change is the major reason why we believe this technology will start to become prominent as it allows the reuse of existing installed cameras," says expert in crowd counting and behaviour analytics Dr Rob Dupre.

Virtual coaches are also one of the most promising technology applications. The user has a device in front of them, deep neural networks receive the stream from the device, process it and provide information on a joint's localisation. There are additional algorithms that determine what exercise was completed, how many times the user did it and if there were any mistakes like hips going higher than shoulders when doing a plank, and so on.

Elsewhere, an advanced tracker has been designed for the retail market. It can be used for counting at entrances and exits, using other business intelligence, or BI, metadata such as how long someone has been queuing, whether there are more than a set number of people waiting to be served or just as a way of detecting which parts of the store generate the most interest to customers.

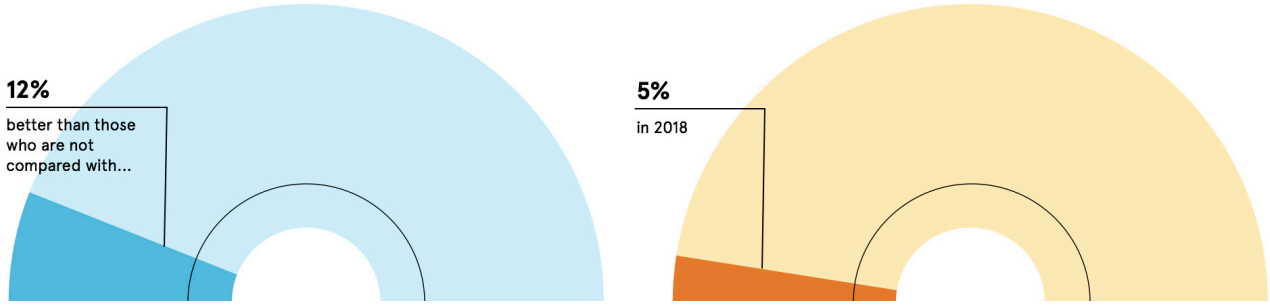
Typical motion-based tracking engines would struggle with false detections caused by moving doors or changes in illumination, but the use of pose estimation eliminates many of these issues.

# Return on AI

## Pose estimation is just one slice of the AI pie, so what can early adopters learn from business users of AI so far?

### Businesses in the UK who were using AI at scale in 2019 were performing...

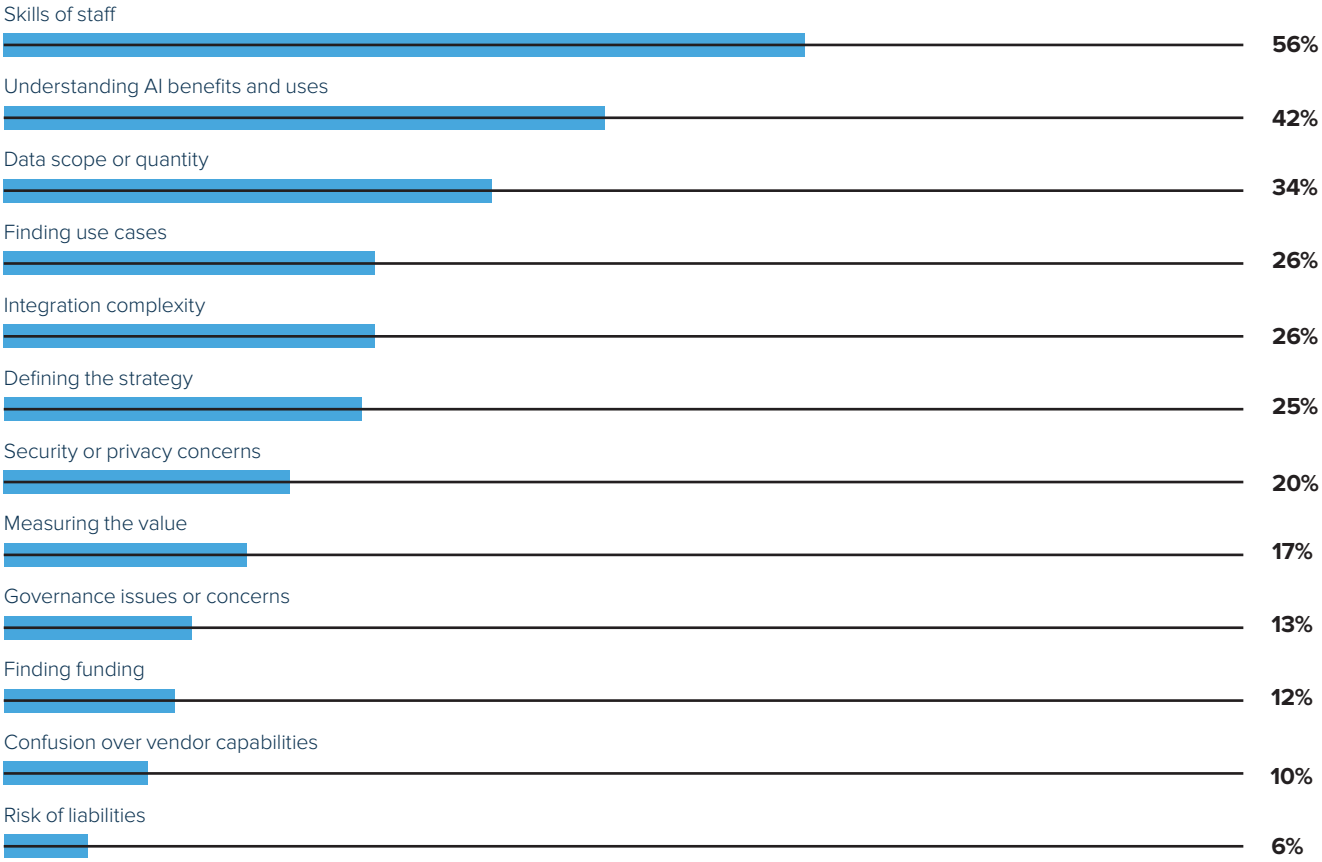
*Microsoft 2019*




### But to ensure organisations are fully leveraging AI technologies, they still need to overcome some key barriers

Top challenges to AI/machine learning adoption as identified by CIOs around the world

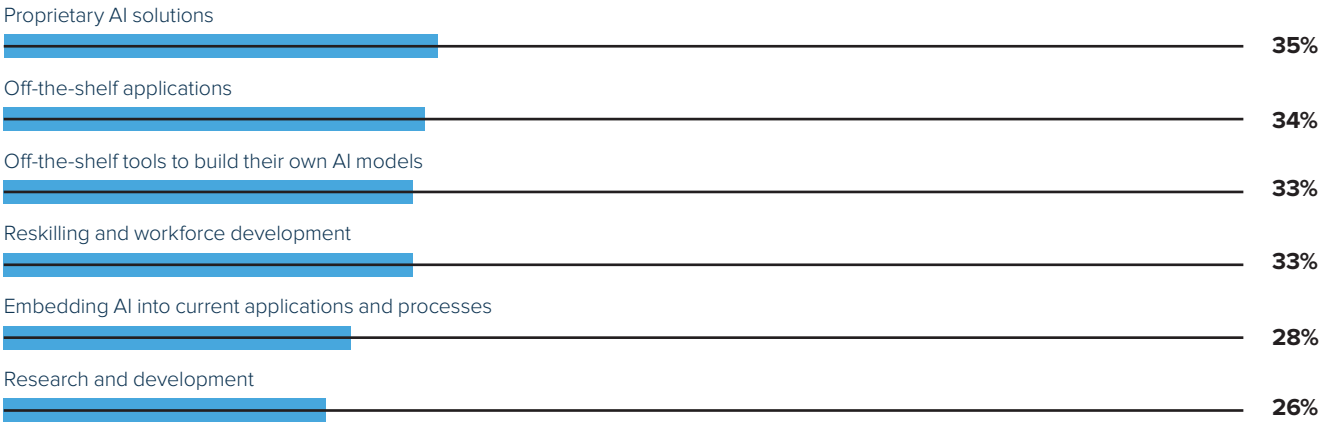
*Gartner 2019*



## How companies are investing in AI

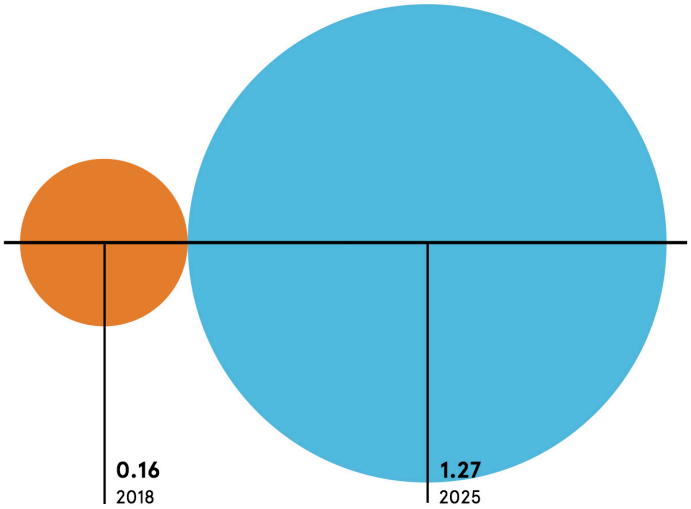
Areas where global companies are planning to heavily invest in all areas of AI over the next 12 months

*IBM 2020*

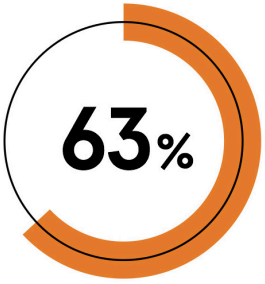



## And spend on computer vision is set to increase over the next 5 years

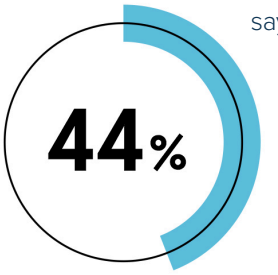
Estimated European market revenue for computer vision ($bn)




## Even though there are some barriers to implementation, organisations are already seeing tangible business benefits of adopting AI




of global executives whose companies have adopted AI say it has provided an uptick in revenue




say AI has reduced costs

# Harnessing the power of human pose estimation

**Barriers to adoption are being overcome so human pose estimation looks set to go mainstream**

**Pavel Akapian, Lead Machine-learning R&D Engineer at InData Labs**

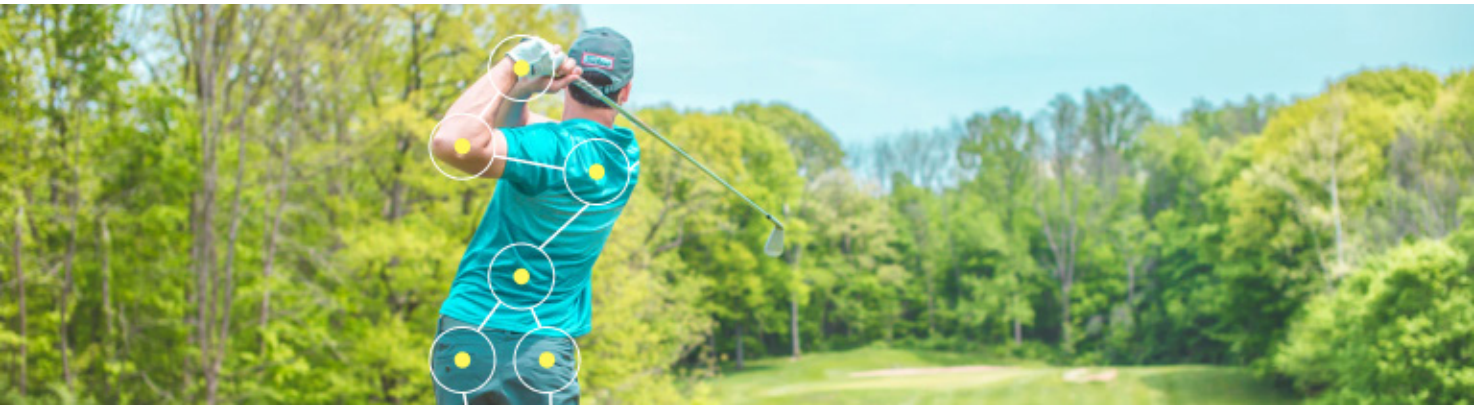
The potential for human pose estimation is immense, with a wide range of applications and use-cases. So why hasn't it been subject to the same level of hype as other kinds of artificial intelligence (AI)?

The answer is likely to be down to human psychology; technologies that evoke excitement and surprise generate the most hype. When it comes to AI, people have strong feelings, either for or against it. This is because technologies that demonstrate machines can create something new – write music, poems, draw paintings, create deepfakes, change age or sex – are subject to the greatest levels of publicity.

The same was true to some extent when human pose estimation broke through a few years ago. There was an early buzz among startups around its use in augmented reality, or AR, specifically with applications where users could try on clothes or costumes virtually.

However, the technology hit a roadblock. Its development was expensive and the business value of trying on real clothes is hard to define. Indeed, creating try-on clothes apps remains a difficult and costly project, as it involves not only human pose estimation development but also 3D clothes modelling and rendering that will look natural on a person in motion.

# Challenges to adoption



There are other challenges to the adoption of human pose estimation. The first is that publicly available neural networks work well only on a limited number of simple poses. Analysing fitness or any other kind of sports takes exact accuracy of the poses that are unnatural for a human when they are at home or walking the streets.

This hurdle can be removed by collecting a dataset for necessary poses and retraining the neural net using this data. To do this it takes machine-learning research and development (R&D) engineers months of work. So, since the neural network training infrastructure is rarely in open access, it should be developed from scratch according to the business case.

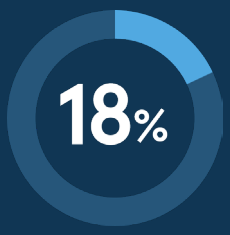
The second hurdle is that one human pose estimation is not enough. It takes some extra work by data scientists to create physical activity type recognition algorithms, repetitions counting and error analysis.

Another important thing for improving accuracy is the post-processing algorithms that significantly decrease neural network errors for human pose estimation. For example, they decrease keypoint trembling or remove confusion about mixing the right and left legs. Unfortunately, there's no common approach to complete these tasks yet. At InData Labs this kind of algorithm is developed according to the client's needs.

However, both these problems are easily solved by hiring the right team of data scientists and machine-learning R&D engineers at the product development stage. The data science community and the market, in general, are evolving rapidly and it's not hard to find professionals.

A third problem often occurs during the pure engineering development stage, when we need to port algorithmic solutions to the user's device, usually a mobile device. It's difficult to find software engineers with experience in porting neural networks and developing custom mathematical algorithms for certain devices. Even a person with multi-level expertise in mobile app development, and who is versed in maths algorithms, can face problems as this kind of work takes hands-on experience.

*McKinsey 2018*




of organisations say they have a clear strategy in place for accessing and acquiring the data that enables AI work

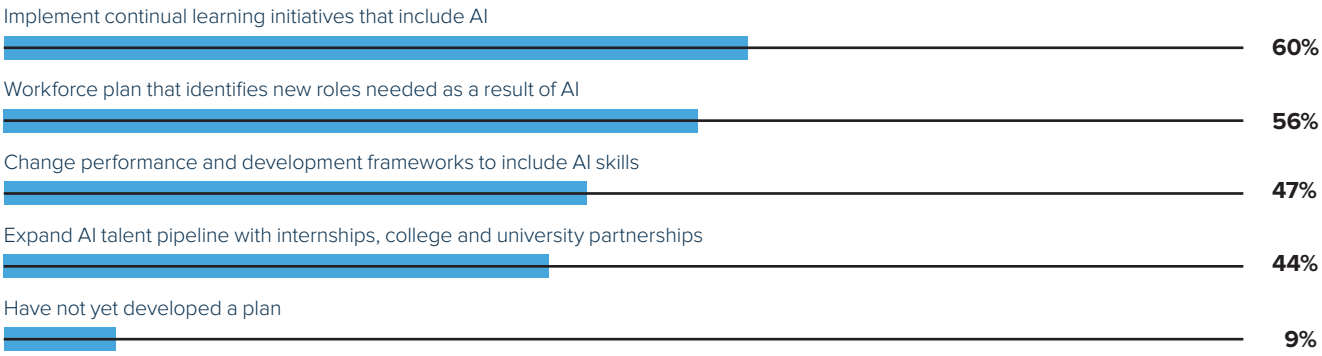
# Huge potential

Despite these stumbling blocks, human pose estimation is one of the most mature AI technologies. Considering its ripeness and adaptability, it only gives way to facial recognition technology. Yet it's still only at the early stages of its potential.

## Building an AI-ready workforce

Percentage of technology executives who plan to do the following

*PwC 2019*




The technology currently sits in the third Gartner hype cycle stage, the slope of enlightenment. After the disillusionment of virtual try-on clothes apps, several startups have begun using the tech in projects that are easier to monetise, such as virtual fitness and yoga coaches.

Virtual coaches can streamline access to professional workouts and help people improve their health. This has never been more relevant than it is currently, with the coronavirus pandemic keeping people away from gyms. Human pose estimation can help people improve their health and get through these hard times.

Elsewhere, organisers of major sports events can use the technology to create more engaging content during broadcasting. Human pose analysis can also help professional teams polish their moves without expensive wearable sensor systems. Meanwhile, thanks to advanced customer behaviour analysis, shops can minimise thefts and shoplifting. Socials can use this technology to provide users with fun and viral content.

# Bright future

**Those who recognise the potential of human pose estimation have a bright future**

At InData Labs we believe that in the future, the development of human pose estimation could replace motion capture processes. By replacing pricey and complex systems like Kinect with a cheap camera, there is potential for human pose estimation within the video production or gaming industry. Those who recognise the potential of human pose estimation have a bright future. On the one hand, the business value is clear. On the other, we have a combination of knowledge, experience and the right technologies to develop projects with predictable results and budget. Considering the technology is ripe, human pose estimation is destined to gain further momentum over the next five years.

This means underestimating human pose estimation's impact and potential will soon fade away, and we will see it widely used in everyday life and business.

# How is human pose estimation used across different sectors?

**As the technology continues to develop, human pose estimation is finding new applications across sectors**

Can you do the Cool Ranch Dance? Your own moves are between you and the bedroom mirror but, if you were paying attention to this year's Super Bowl ads, you'll know that once you've uploaded them, you can dance like a professional dancer. Or “AI dance” at least with smooth footwork transposed onto your body through the magic of human pose estimation.

The technology is poised to impact industries, notably health and fitness, and entertainment, and the intersection of the two: dancing. Take the app just described. It's called Sway, from Humen.Ai based in Emeryville, California. It uses the tech to track and analyse the skeletal structure of its users and then produces synthetic media, enabling creation of “dancefakes”, or deepfakes of you performing dance moves and more.

"Human pose estimation is an integral part of our AI-driven system," says Humen co-founder and chief executive Tinghui Zhou. Since its Super Bowl Cool Ranch Dance launch advert in January, with Lil Nas X, Sam Elliott and Doritos, Zhou says the team has been aggressively growing their content library in terms of both quantity and diversity beyond dancing.

Humen is building a community of content creators featuring dancers, athletes, musicians and actors, and laying the foundations for its ambitions in sectors including gaming, animation and telepresence, all while

working with Diplo to generate the dance sequences in his recent On My Mind music video.

Building on the Microsoft Kinect heritage, some virtual and augmented reality gaming-focused systems are also beginning to track humans interacting with both real-world and virtual objects, which rely on recreating human movement accurately in game worlds.

Related fields will benefit and we're at the point where Snapchat has added basic pose estimation to its Lens Studio. In fact, any creative field that involves creating and animating 3D human models is likely to adopt human pose estimation technology.
"3D pose estimation is starting to democratise animation and a lot of major fields rely on animation, such as gaming, virtual influencers and CGI [computer-generated imagery] in movies," says Sudarshan Chandra Babu, a deep-learning research engineer at ViGIL Lab at the Indian Institute of Technology in Bombay.

"Fundamental animation workflows are changing. Right now if you want to create an animation it's really hard, but what if you can animate it using a video you shot on your iPhone? You dance and the virtual avatar dances."

> Any creative field that involves creating and animating 3D human models is likely to adopt human pose estimation technology

# Lockdown

On the fitness side, lockdown-friendly artificial intelligence (AI) home workout trainers, such as Onyx, can count reps and correct your form. "We've reached a point where we have fast and accurate 3D pose estimation that provides a great experience in body-weight fitness. But there is still additional work that can be done with complex poses and occlusive scenes," says James Sha, co-founder and chief technology officer at Onyx.

For instance, complex yoga poses can lead to large portions of the body being hidden from view, which requires predicting how the body is actually positioned. "In these instances, the best is yet to come," says Sha.

In professional sports, software such as HomeCourt, which also has a consumer-facing mobile app with a skill-rating system, has been used by NBA basketball teams for performance optimisation and game analytics since summer 2019. There's a similar opportunity for Mixed Martial Arts and the Ultimate Fighting Championship.

There's also much potential for consumer applications in next-generation smart assistants, robots and connected home appliances. "We need human pose estimation in different fields, such as robot-to-human handovers, healthcare robots and kitchen or cooking robots," says Adam Grzywaczewski, senior deep-learning data scientist at NVIDIA. "Think about being able to direct your robot hoover to 'clean there' with a simple voice command."

This extension of simple verbal communication could play an important role in how we can interact with others via future medical technologies and assistive systems. Babu points to work being done in physiotherapy where startups are experimenting with smartphone apps that are able to analyse and correct movement as a human physiotherapist would.

# Outlook

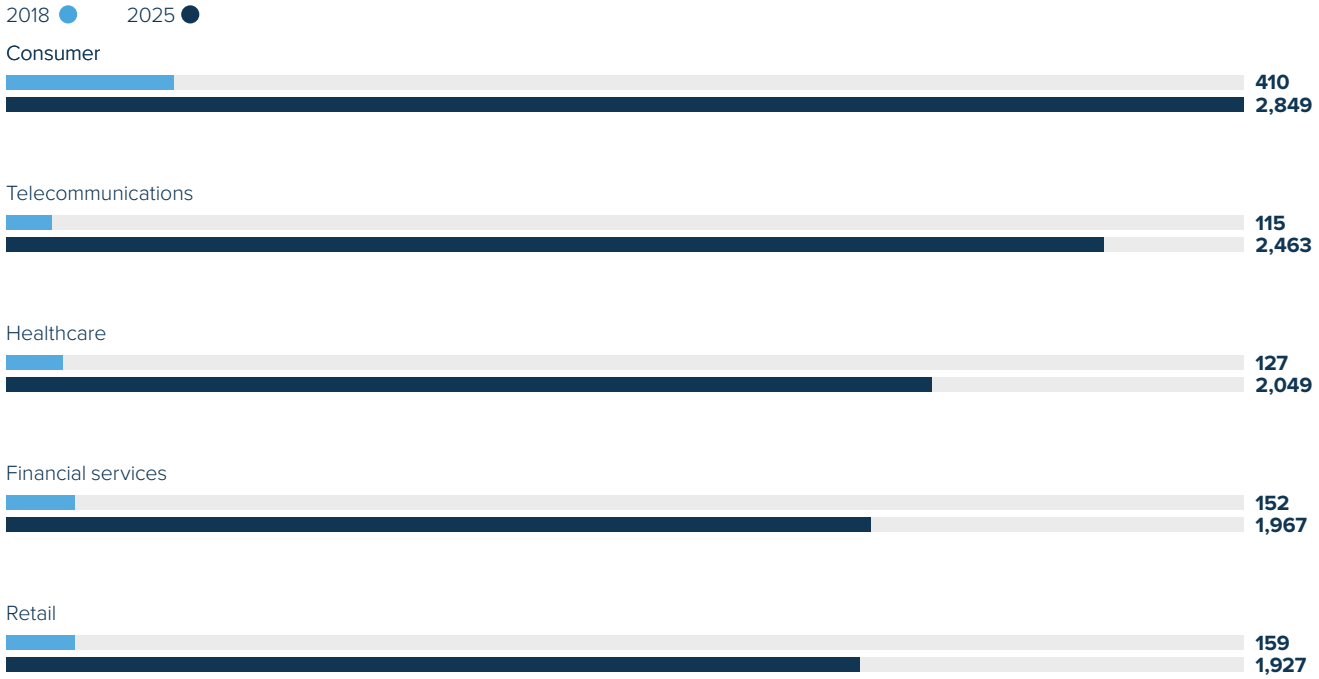
So is the field set to grow or have we reached the peak of what the technology can achieve? This depends on what exactly we're referring to. "2D human pose estimation has almost reached the peak. The AI models are really, really good, but the applications are limited," says Babu. “However, 3D pose estimation from a simple image of video still has a way to go. There has been good progress in research over the past couple of years, so I'm expecting this trend to continue and 3D pose estimation to get really good by, optimistically, 2023.”

There are a number of reasons for the lag in the startup and business side. Firstly, most 3D human pose estimation models work well on academic datasets, but tend not to generalise very well with real-world data. Also, whereas AI models, such as classification and segmentation, benefit from good-quality, large datasets that are relatively cheap to source, 3D pose estimation relies on 3D data with, for instance, more expensive motion-capture setups.

With robust computer vision models available, NVIDIA's Grzywaczewski suggests a counter to problems concerning the financial viability of data processing to the correct level of performance. "New research into semi or self-supervised learning could decrease the cost of data processing considerably, enabling further breakthroughs in pose estimation and computer vision as a whole,” he says.

## The five biggest industries for AI software

Top industries for estimated European market revenue by 2025 ($m)



As part of this effort to improve cost efficiency, and thus open up innovation in the field, NVIDIA is developing AI accelerators that are smaller and have reduced power consumption per computation unit. Its latest A100 graphics processing unit, for instance, is 20 times faster than its predecessor, which makes it possible to develop larger models.

Other key information gaps and missed opportunities include a lack of people who know how to use the technology. This is a result of human pose estimation being less democratised than other tech, but the number of researchers is growing and there are training platforms, such as the NVIDIA Deep Learning Institute, trying to address the problem.

The likes of Google, Facebook and Amazon have also historically not shown interest in funding human pose estimation to the extent of other areas in AI. "But things seem to be changing with virtual reality gaming at Facebook and Google, and Amazon pushing out a few interesting papers," says Babu. "It looks like they're increasing funding."

# Key takeaways

**What are the key things you need to know about human pose estimation? Here are the top five takeaways**

**1.** Human pose estimation refers to a number of computer vision techniques concerned with determining the joints, known as key points, on the human body – knees, hips, elbows, neck, shoulders, feet – in both 2D images and 3D video. This enables a number of estimates, including the body's position, such as lying down or stretching, its location within a scene, its movement and even the ability to infer the activity a body is performing, such as yoga, basketball or dancing.

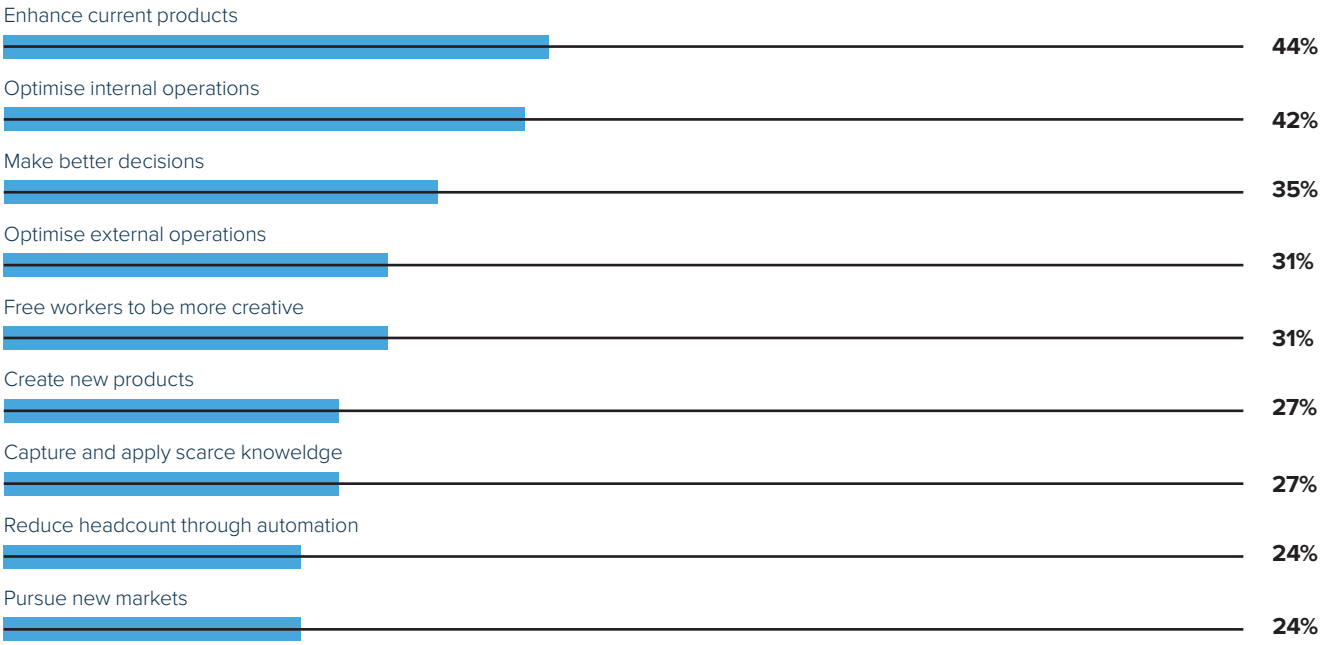
**2.** The top use-cases of human pose estimation technology range from performance optimisation in professional sports to transforming animation workflows in gaming and CGI (computer-generated imagery), building on the initial promise of Microsoft Kinect. In consumer applications, apps such as Humen.Ai's artificial intelligence (AI) dance filter Sway and virtual fitness trainer Onyx are demonstrating the capabilities of pose estimation to a wider audience. The potential for human pose estimation goes beyond fitness, virtual reality gaming and entertainment to encompass everything from assistive healthcare and physiotherapy to robot-to-human interaction and business intelligence.

**3.** The business impact in sectors including security and retail has also been accelerated by the current disruption from the coronavirus pandemic. Pose estimation is a machine-learning technology that can automate tasks such as tracking venue capacity, entrance, exit and queue counting, as well as estimating social distancing. It's also compliant with European Union data privacy regulations as it doesn't identify personal features, just the joints on an anonymous body.

**4.** The main challenges to adoption of pose estimation include concerns that apply to AI in the wider sense, such as trust in the technology, and finding suitable engineering talent and funding from big tech companies for research. There are also specific hurdles including the relatively high cost of building large, high-quality datasets, compared to other machine-learning models, and the problem of applying theoretical models to real-world data to be useful and commercially viable.

## Leading benefits of AI adoption

Percentage of US IT leaders and business executives who rated the following as a top-three benefit *Deloitte 2018*




**5.** While 2D human pose estimation has reached something of a peak, 3D human pose estimation is at the beginning of its development, and needs funding and talent to further innovation. With research and commercial interest from Facebook, Google, Amazon, Snap and others, together with AI accelerators and training platforms from NVIDIA, as well as being a natural fit with emerging technologies such as augmented reality, human pose estimation could be a key component in bringing together the physical and digital worlds.

## About InData Labs

InData Labs is a leading data science firm and AI-powered solutions provider with its own R&D center. Having a mission to bring the power of AI to every business, we help organizations of any size create intelligent products and services or shape intelligent business processes.

Since 2014, our solutions and consulting services help our clients to get valuable insights into data, automate repetitive tasks, enhance performance, add AI-driven features, and prevent cost overruns.

**Cyprus**
16, Kyriakou Matsi,
Eagle House,
Agioi Omologites,
Nicosia

**Lithuania**
Konstitucijos av., 21A,
Vilnius, Republic of Lithuania,
LT-08130

**Singapore**
6 Shenton Way,
OUE Downtown 2,
Singapore

linkedin.com/company/indata-labs
facebook.com/indatalabs
twitter.com/indatalabs

**indatalabs.com**

RACONTEUR

InData Labs