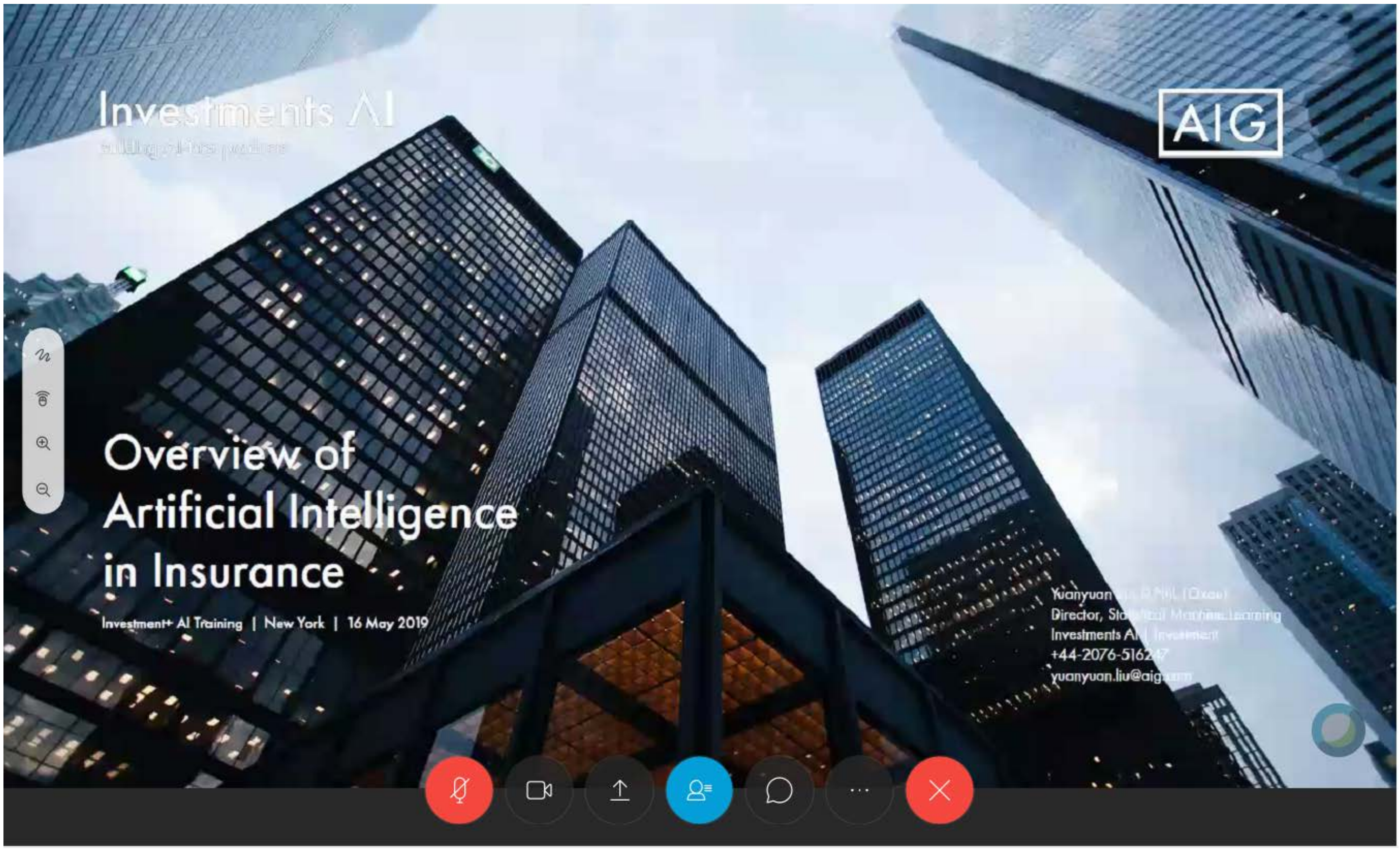


Looking for disruptive opportunities

Overview of AI

Investments AI

Highly confidential – Do not copy

AIG

Why AI?

"Over time, we will move from mobile-first to an AI-first world"

Sundar Pichai, CEO of Google

"Our responsibility is to have AI augment the human ingenuity and the human opportunity"

Satya Nadella, CEO of Microsoft

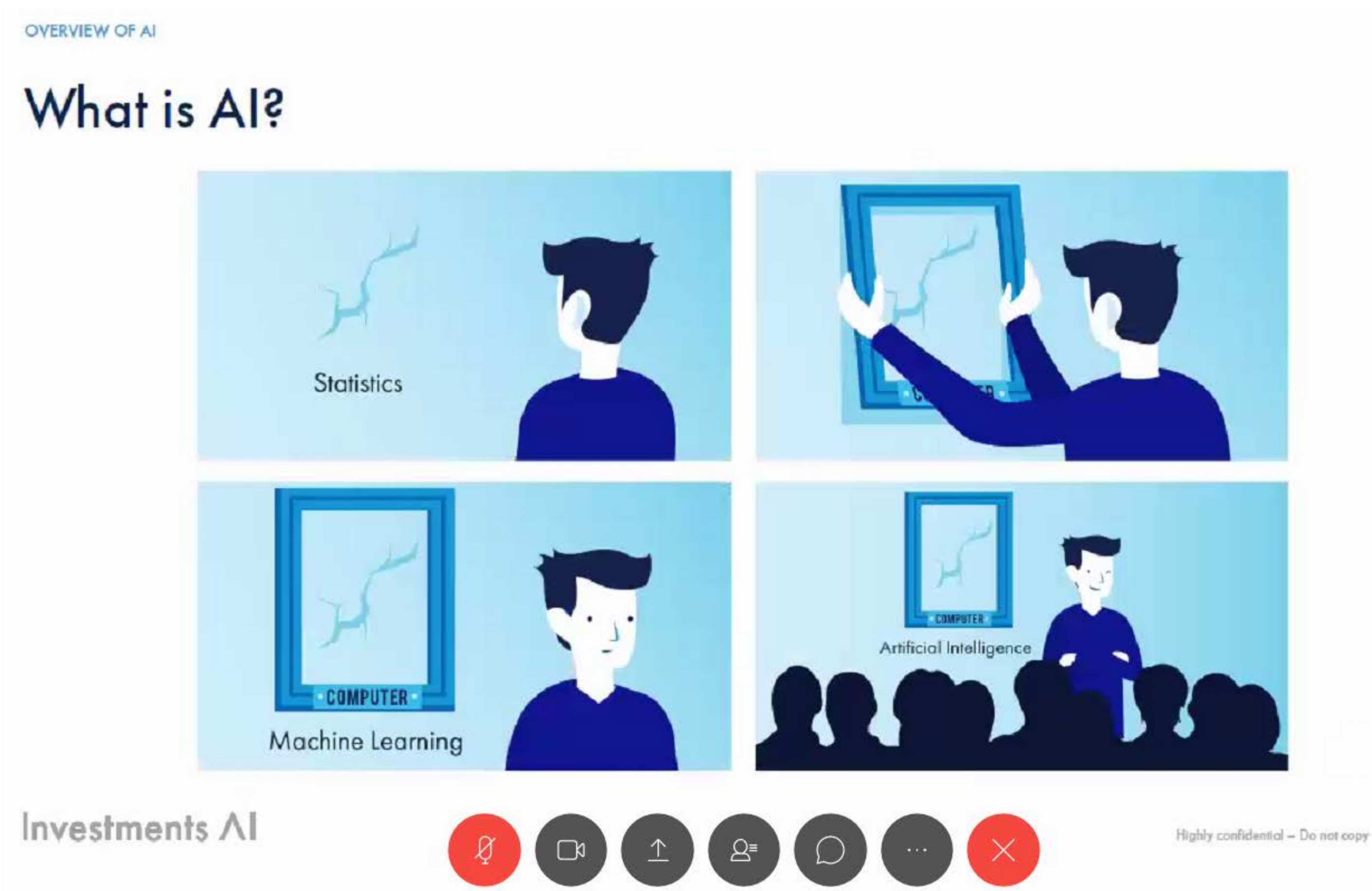
"I believe that at the end of the century, general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted"

Alan Turing, Father of General-Purpose Computer

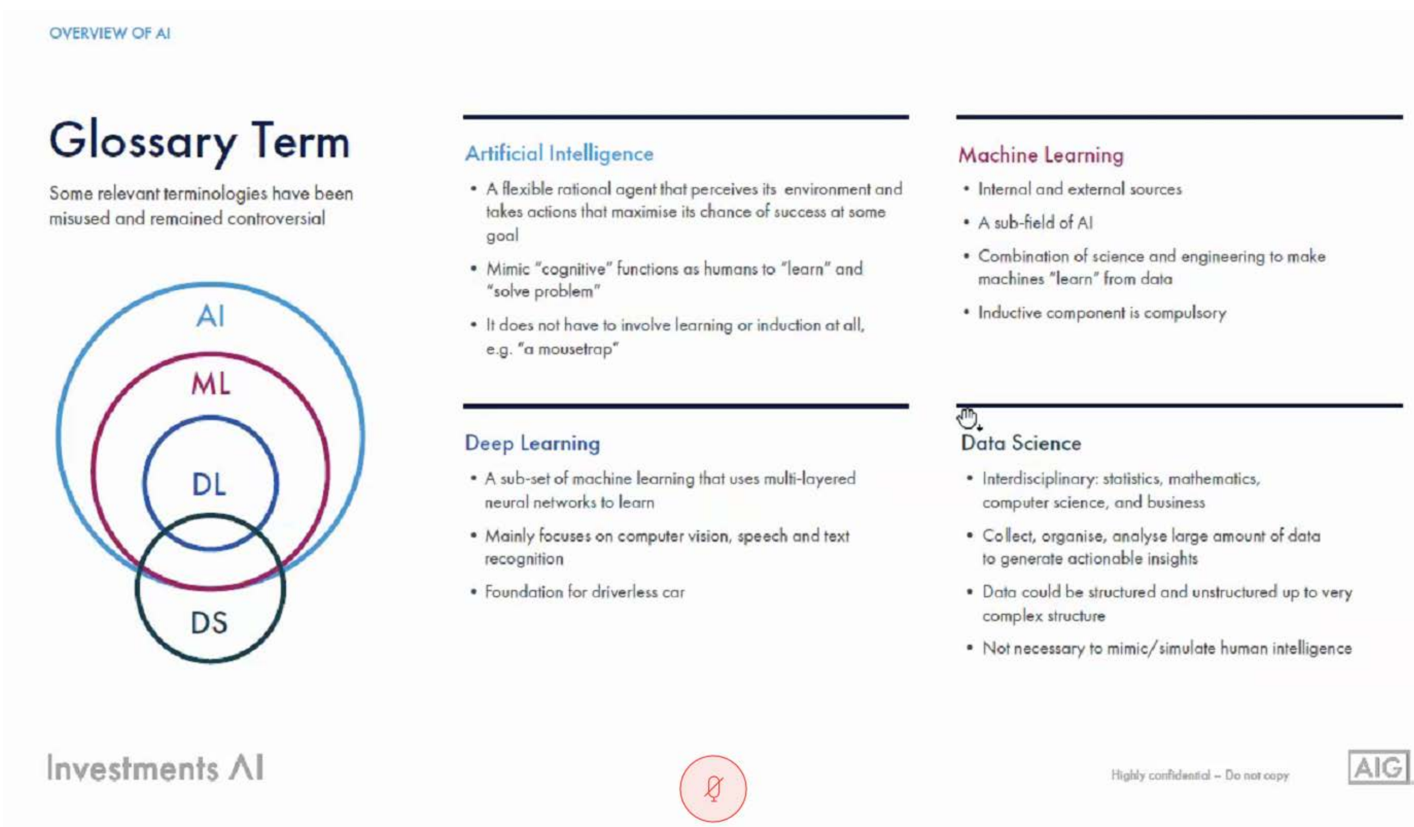
Investments AI

Highly confidential – Do not copy

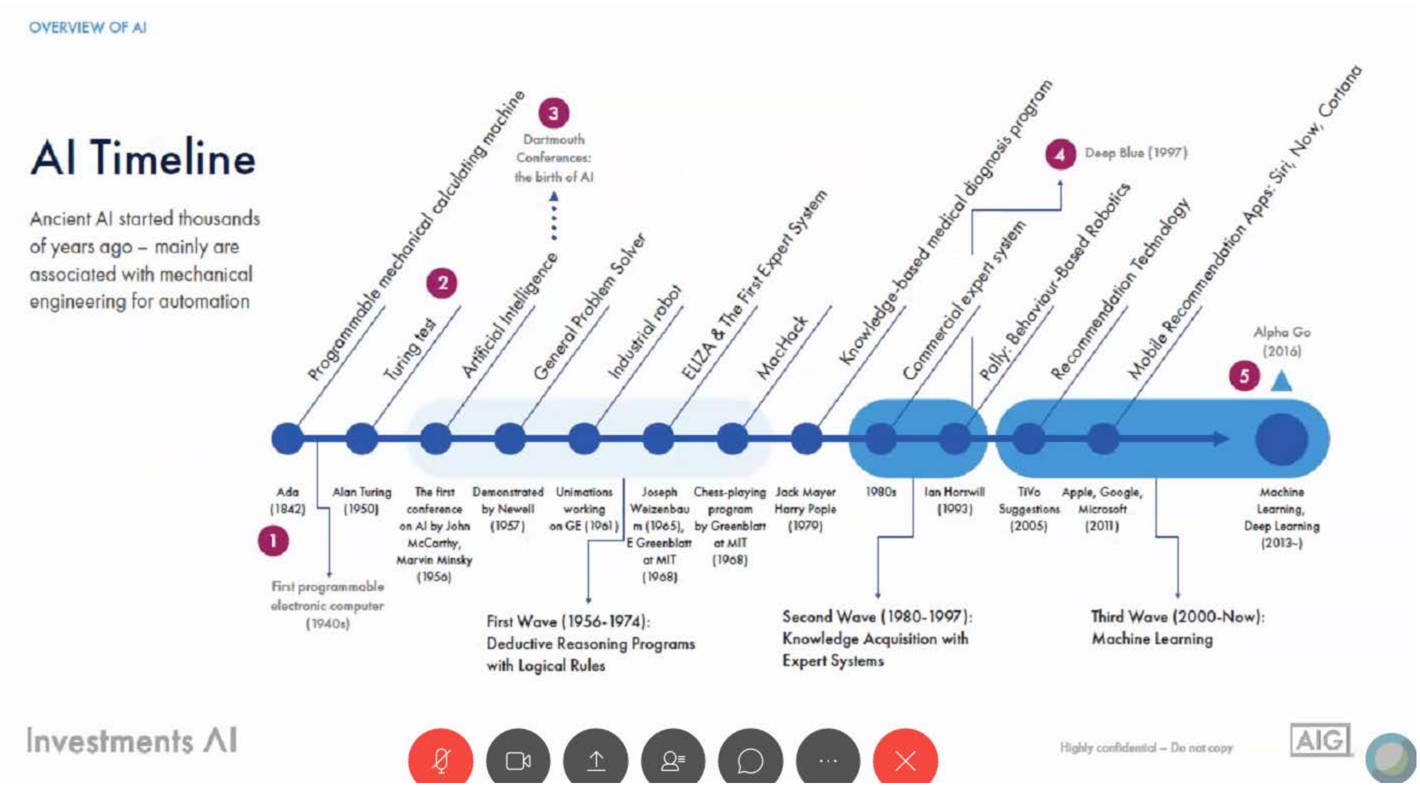
AIG



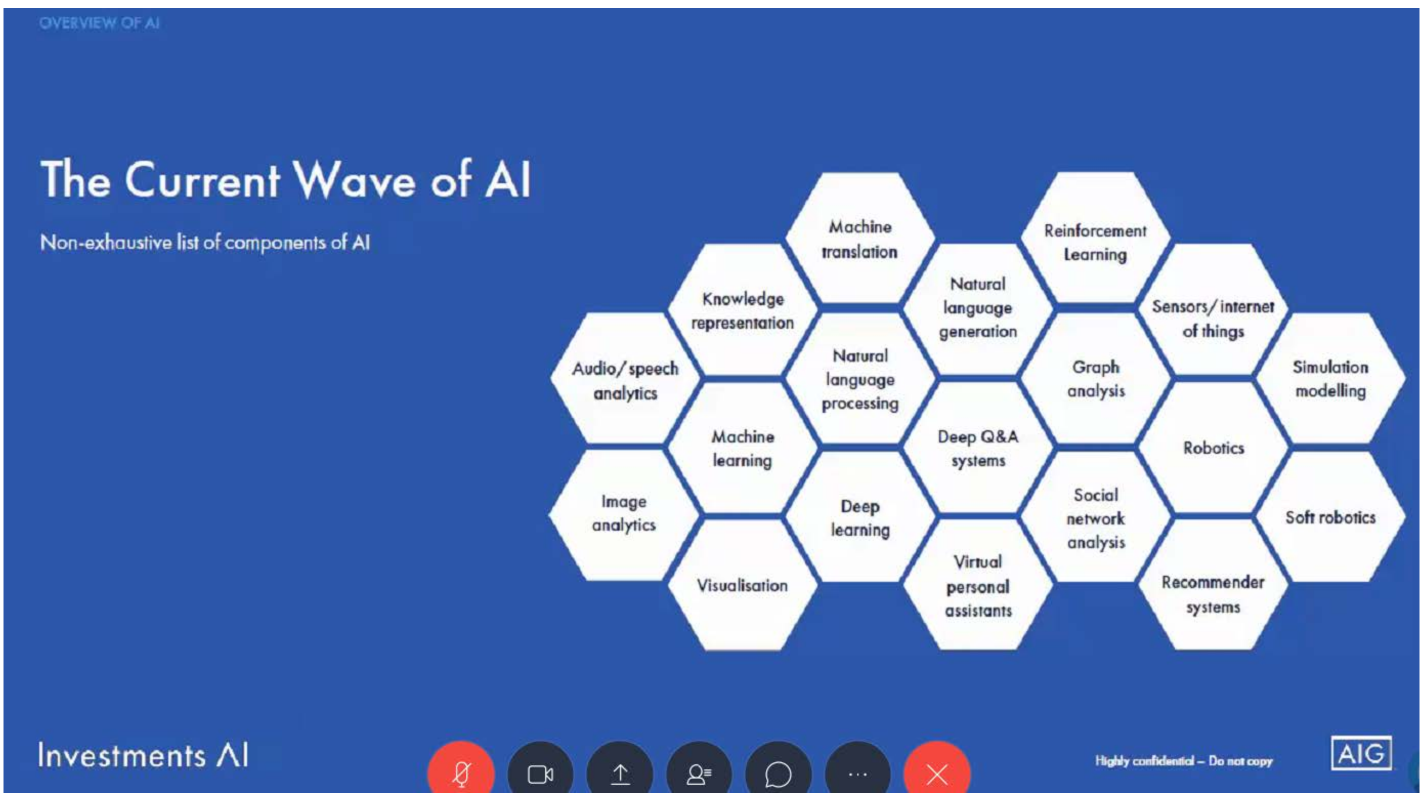
AI = statistics -> Machine Learning -> AI

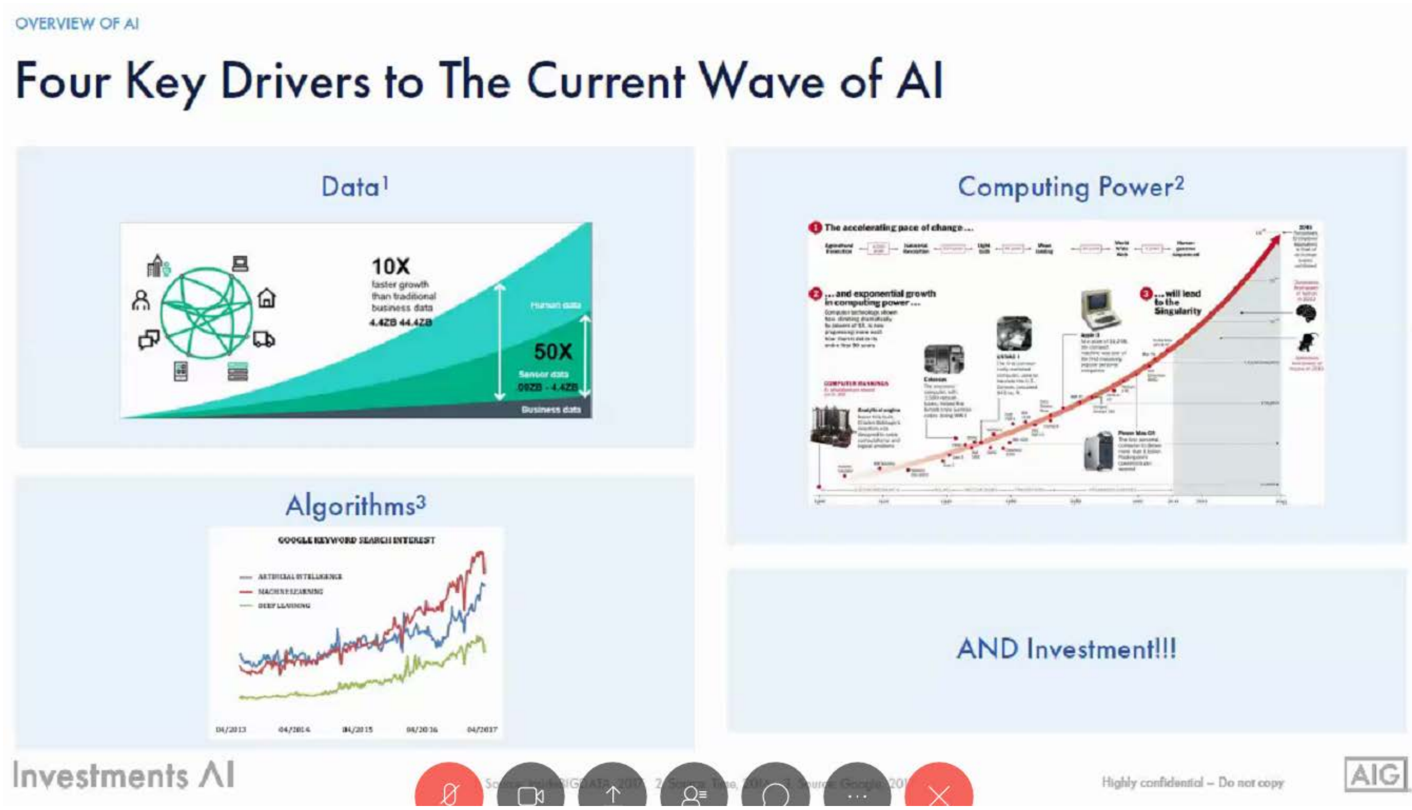


More information can be found at AIG AI Contact Page: https://contact.aig.net/pc/scn/Pages/Science%20Landing%20Page.aspx



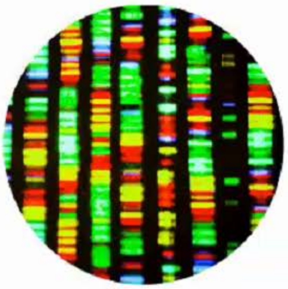
Third wave starting 2000 and ending now – Noway Ai = Deep Learning





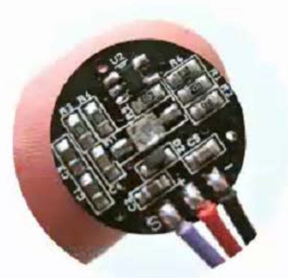
Exponential growth of data, computing power and algorithms

New Data Sources Available to Fuel Insight Creation



Genome sequencing

Massive decreases in cost of genome sequencing, from $96M in 2001 to $4K today



World of sensors

Trillions of internet-connected sensors collecting and sharing data to improve insight

Narrative Science

Natural language written by machines to make data simple and easier to understand

Investments in AI

US-based companies absorbed 66% of all AI investment in 2016, followed by China with 17%

The current AI wave is poised to finally break through

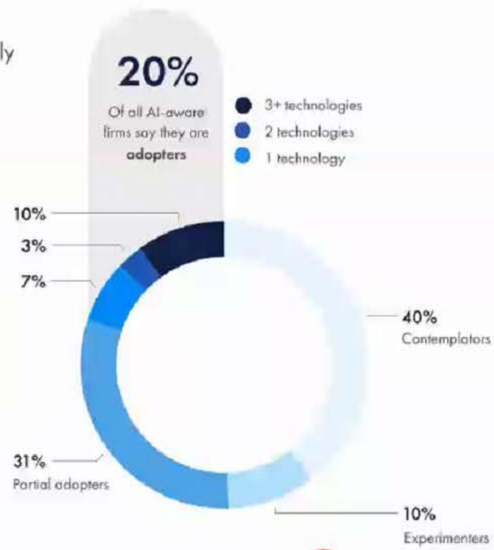
Investment in AI is growing at a high rate, but adoption in 2017 remains low:

In 2016, companies invested
$26bn - $39bn
In artificial intelligence

Tech giants $20bn - £30bn

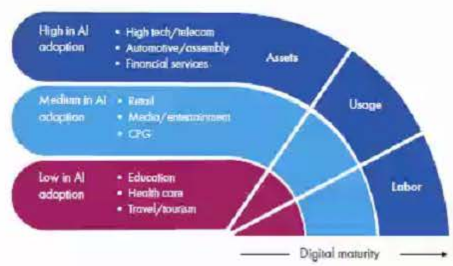
Startups $6bn - £9bn

3x External investment growth since 2013



How companies are adopting AI

Investment in AI is growing at a high rate, but adoption in 2017 remains low



Six characteristics of early AI adopters

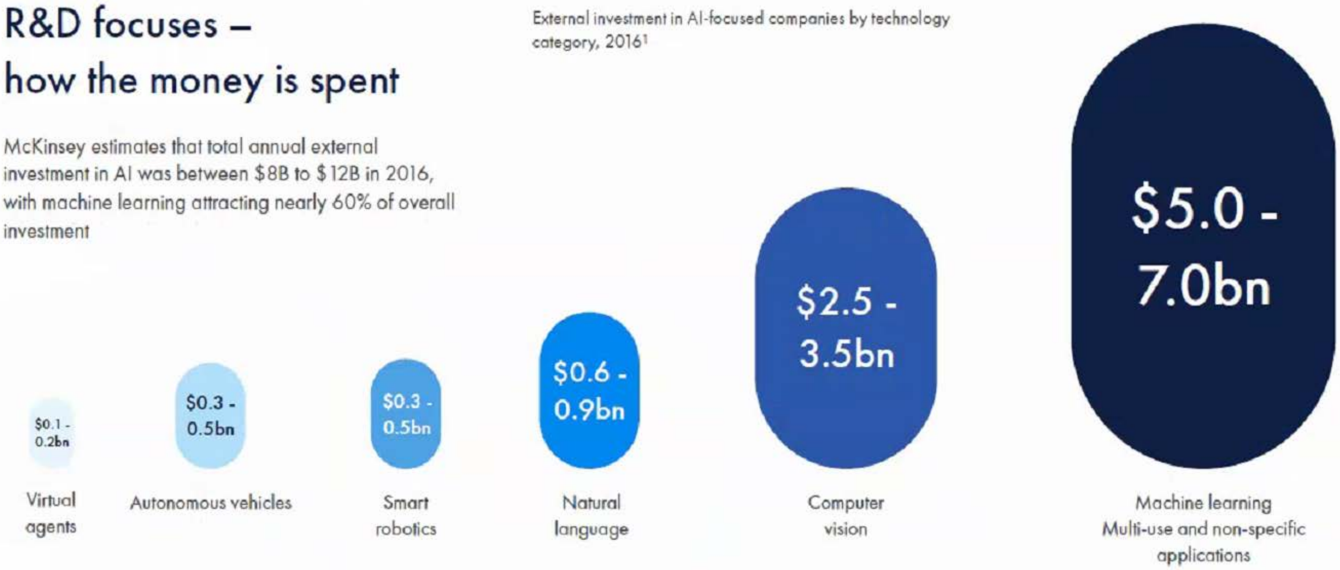
- Digitally mature
- Larger business
- Adopt AI in core activities
- Adopt AI multiple technologies
- Focus on growth over savings
- C-level support for AI

R&D focuses – how the money is spent

McKinsey estimates that total annual external investment in AI was between $8B to $12B in 2016, with machine learning attracting nearly 60% of overall investment

Machine learning received the most investment, although boundaries between technologies are not clear-cut

External investment in AI-focused companies by technology category, 2016[1]



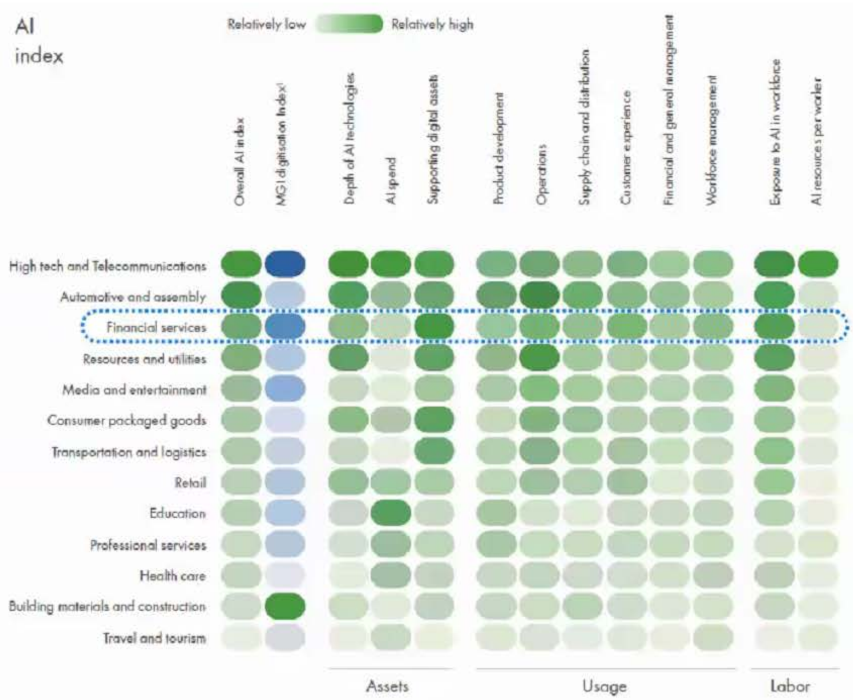
[1] Estimates consist of [illegible] VC investment [illegible] AI-focused [illegible]. PE inv[illegible] in AI-relat[illegible] and [illegible] corpor[illegible] includes only disclosed data available in d[illegible]

Future AI demand

Financial Services are leading the digitalisation, supporting digital assets, exposure to AI in workforce; but relatively weak at AI spend, and AI resources per worker

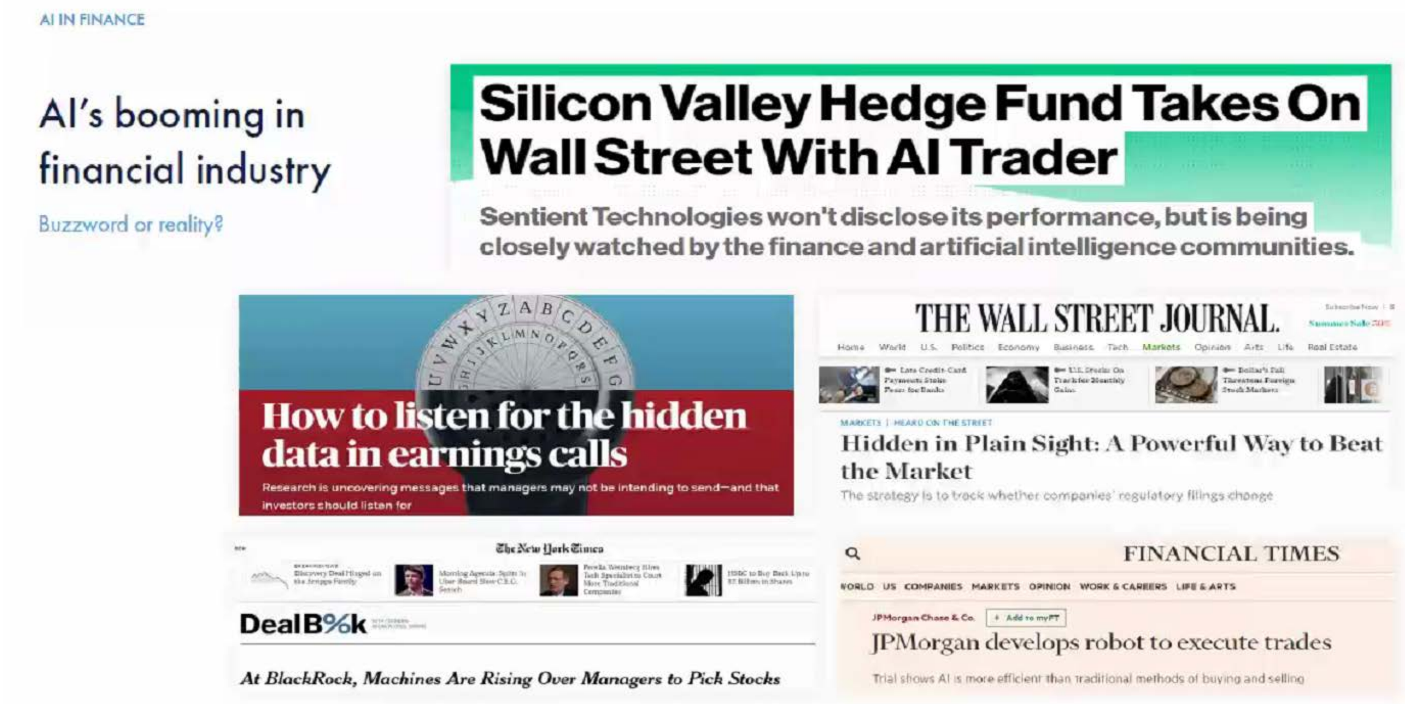
AI adoption is occurring faster in more digitised sectors and across the value chain

AI index



Investments/Finance is catching up on AI demand

No differences between non-adapters and partial adapters -



Expected explosive interest in AI in Finance

AI IN INSURANCE

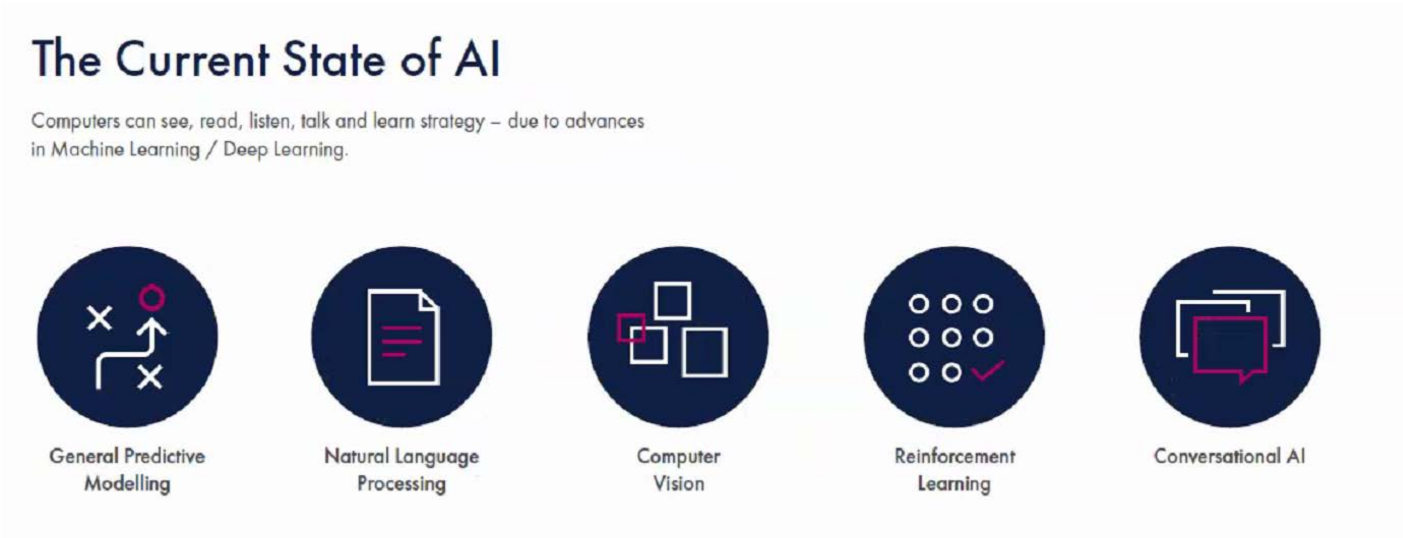
What is Insurance?

At a high level, excellence in insurance can be associated to success is in its key pillars

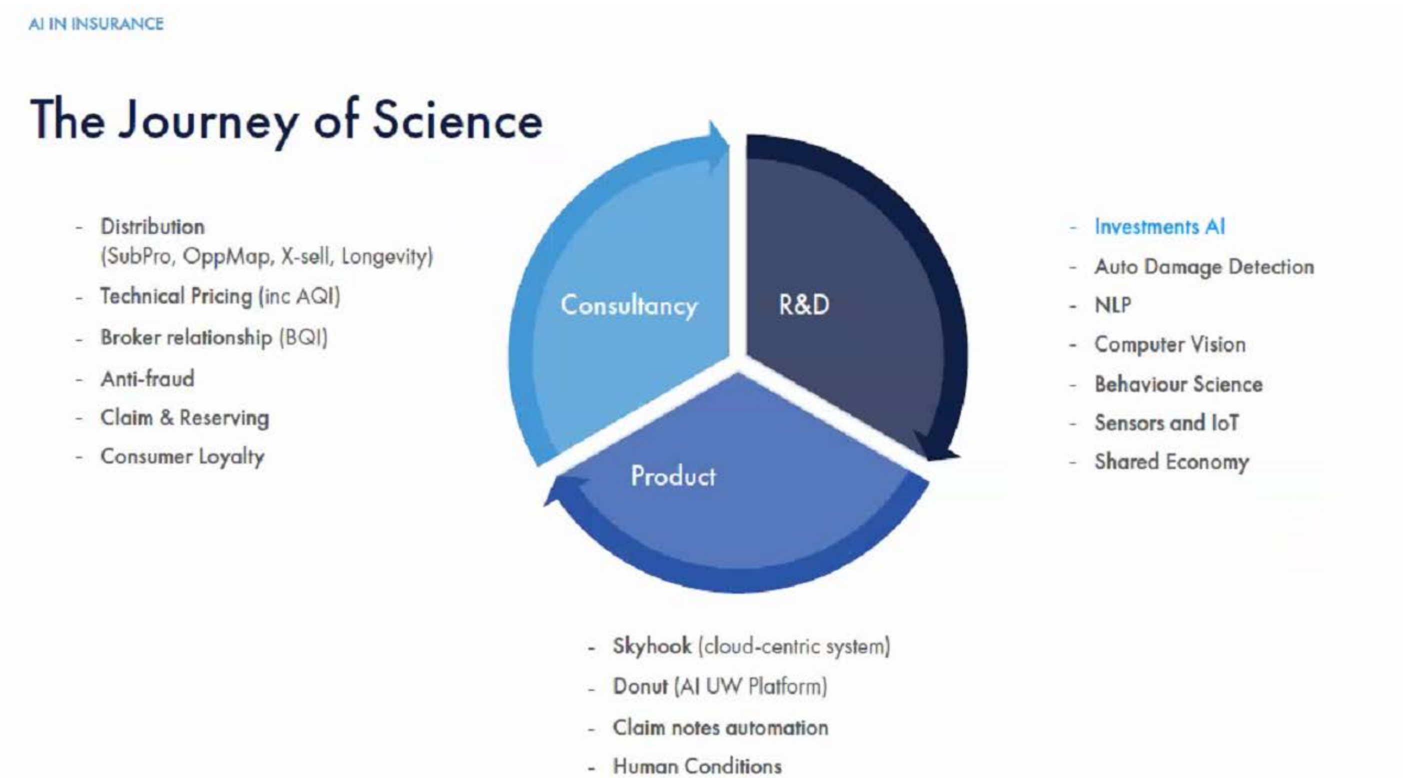
AI IN INSURANCE

The Nine Killer Applications of Digital Technology in General Insurance

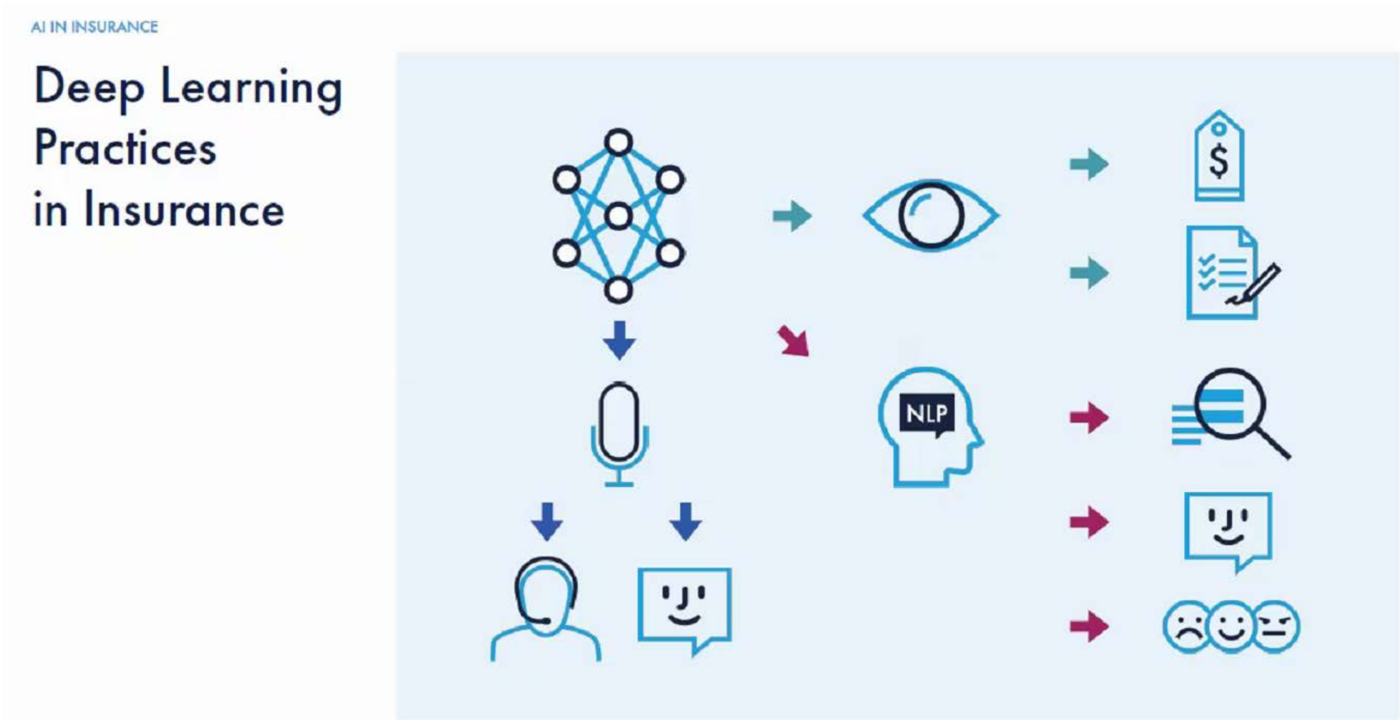
Deloitte – Cybner risk can be huge



Most mature: genreal predictive modelling



Consultancy – Advise business on AI R&D AWS centric

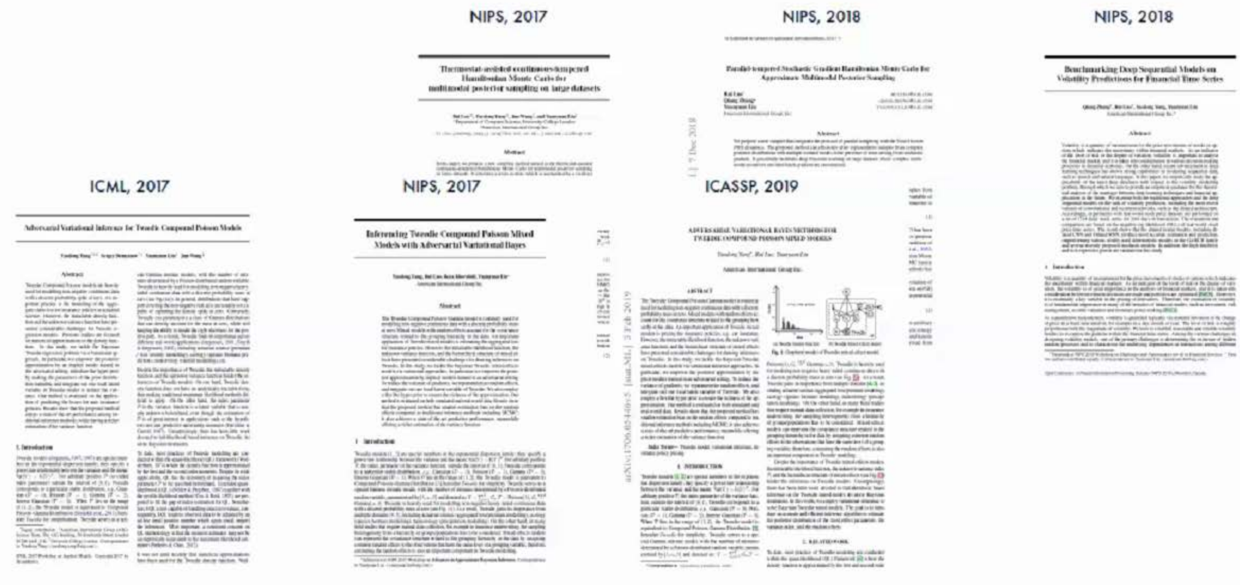


“Sharp Eye” scan your home to determine what needs be insured

Finance text data: contracts, claims, coverages – sentiments

Research

Recent Publications



Research Case I – Bayesian Tweedie

Tweedie Distribution

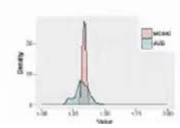
Tweedie Compound Poisson models are heavily used for modelling non-negative continuous data with a discrete probability spike at zero. An important practice is the modelling of the aggregate claim loss for insurance policies in actuarial science. However, the intractable density function and the unknown variance function have presented considerable challenges for Tweedie regression models.

Variational Inference

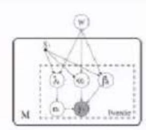
We tackle the Bayesian Tweedie regression problem via a Variational approach. In particular, we empower the posterior approximation by an implicit model trained in the adversarial setting, introduce the hyper prior by making the parameters of the prior distribution trainable, and integrate out one local latent variable in Tweedie model to reduce the variance.

Why we need it?

Our method is evaluated on the application of predicting the losses for auto insurance policies. Results show that the proposed method enjoys a state-of-the-art performance among traditional inference methods, while having a richer estimation of the variance function.



Posterior distribution of the index parameter P for Adversarial Variational Bayes (AVB) is much more flexible than the distribution learned using MCMC method. This allows to more accurately model the loss distribution.



Model learns a distribution of global weights W, which combined with the client features X to infer the Tweedie distribution parameters λ, α and β. Then loss Y can be sampled to obtain the average and all quantiles.

Research Case I – Bayesian Tweedie

Results: AVB method shows state-of-the-art performance, even compared with the TDBoost method which is considered to be the strongest.

Table 1: The pairwise Gini index comparison with standard error based on 20 random splits

Baseline / Model	GLM	PQL	Laplace	AGQ	MCMC	TDBoost	**AVB**
GLM	/	$-2.97_{6.28}$	$1.75_{5.68}$	$1.75_{5.68}$	$-15.02_{7.06}$	$1.61_{6.32}$	$\mathbf{9.84_{5.80}}$
PQL	$7.37_{5.67}$	/	$7.50_{6.26}$	$7.50_{5.72}$	$6.73_{5.95}$	$0.81_{6.22}$	$\mathbf{9.02_{6.07}}$
Laplace	$2.10_{4.52}$	$-1.00_{5.94}$	/	$8.84_{5.36}$	$4.00_{4.61}$	$21.45_{4.84}$	$\mathbf{20.61_{4.54}}$
AGQ	$2.10_{4.52}$	$-1.00_{5.94}$	$8.84_{5.36}$	/	$4.00_{4.61}$	$21.45_{4.84}$	$\mathbf{20.61_{4.54}}$
MCMC	$14.75_{6.80}$	$-1.06_{6.41}$	$3.12_{5.99}$	$3.12_{5.99}$	/	$7.82_{5.83}$	$\mathbf{11.88_{5.50}}$
TDBoost	$17.52_{4.80}$	$17.08_{5.36}$	$19.30_{5.19}$	$19.30_{5.19}$	$11.61_{4.58}$	/	$\mathbf{20.30_{4.97}}$
AVB	$-0.17_{4.70}$	$0.049_{5.62}$	$3.41_{4.94}$	$3.41_{4.94}$	$0.86_{4.62}$	$11.49_{4.93}$	/

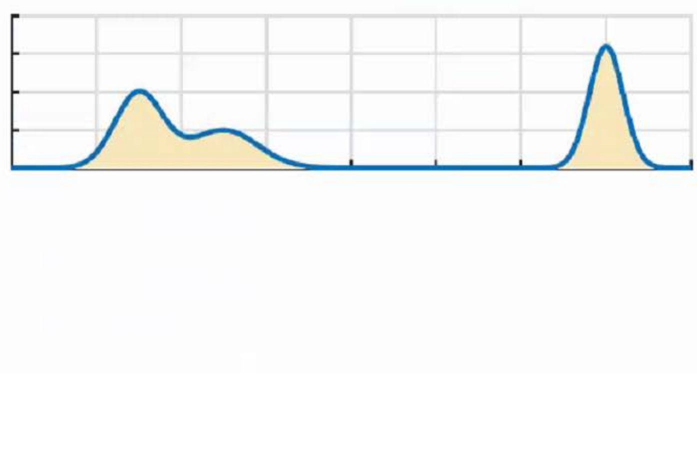
We evaluate our method by predicting losses for auto insurance policies. Dataset properties:

- 10,296 auto insurance policies, splitted on 50% / 25% / 25% for train / validation / test sets,
- 56 features, which describe general information about a driver and its vehicle,
- 61% of drivers have no claims, 9% of drivers make 65% of all claims

Research Case II – TACT-HMC

Motivation

Bayesian learning with MCMC captures the uncertainty of the learned parameters. However, when the distribution of interest contains multiple modes, efficient exploration across all those modes becomes difficult. In particular, when the number of modes is large, the "distant" modes could beyond the reach of any closest modes; this would lead to the so-called pseudo-convergence, where the ergodicity guarantee of MCMC methods breaks.



To make it worse, Bayesian learning on large datasets is typically conducted in an online setting: at each iteration, only a mini-batch of data are used to update the model. While the requirements for computation are substantially reduced, mini-batches introduce noise and hence additional uncertainty into parameters, which makes multimodal posterior sampling even more difficult.

Research Case II – TACT-HMC

Conventional attempt: Hamiltonian Monte Carlo

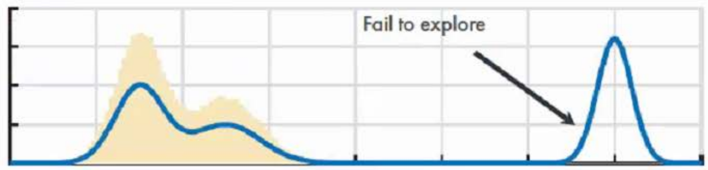
Hamiltonian Monte Carlo is the de facto standard for Bayesian sampling tasks. It first defines a Hamiltonian system upon the potential function of the form

$$U(\boldsymbol{\theta}) = -\log \rho(\boldsymbol{\theta}|\mathscr{D}) = -\log \rho(\boldsymbol{\theta}) - \sum_{i=1}^{N} \log \ell(\boldsymbol{\theta}; \boldsymbol{x}_i) - \text{const}$$

It then simulates the evolution of that system via the Hamiltonian dynamics

$$H(\boldsymbol{\Gamma}) = U(\boldsymbol{\theta}) + \boldsymbol{p}_\theta^\top \boldsymbol{M}_\theta^{-1} \boldsymbol{p}_\theta / 2$$

The Metropolis-Hastings test will be performed at last, serving as a correction for the discretization errors. In the full-batch setting, HMC ends up in a biased histogram, which fails to explore some of modes.



Research Case II – TACT-HMC

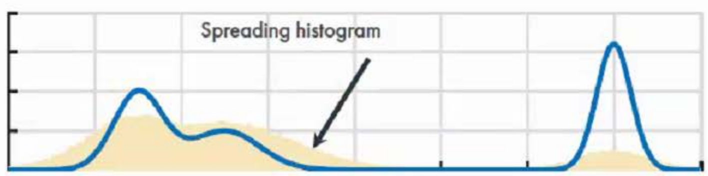
Ingredient I: Continuous Tempering

In physics, continuous tempering is currently a state-of-the-art method to accelerate the exploration of complex free energy surface by means of continuously and systematically varying the temperature of a physical system. It extends the original Hamiltonian system in a way that

$$H(\boldsymbol{\Gamma}) = \lambda(\xi) U(\boldsymbol{\theta}) + W(\xi) + \boldsymbol{p}_\theta^\top \boldsymbol{M}_\theta^{-1} \boldsymbol{p}_\theta / 2 + p_\xi^2 / 2m_\xi$$

By simulating the Hamiltonian dynamics, the extended system travels through the entire phase place.

However, when moving to the mini-batch setting, the noise by mini-batching perturbs the trajectory, leading to a spread histogram



RESEARCH

Research Case II – TACT-HMC

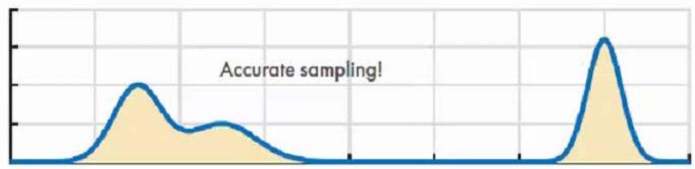
Algorithm

Algorithm 1 Thermostat-assisted continuously-tempering Hamiltonian Monte Carlo

Input: stepsize η_θ, η_ξ; noise level c_θ, c_ξ; (rescaled) thermal inertia $\gamma_\theta, \gamma_\xi$; unit interval of simulation K_0

$\boldsymbol{r}_\theta \sim \mathcal{N}(\boldsymbol{0}, \eta_\theta \boldsymbol{I})$ and $r_\xi \sim \mathcal{N}(0, \eta_\xi)$; $(z_\theta, z_\xi) \leftarrow (c_\theta, c_\xi)$
INITIALISE($\boldsymbol{\theta}, \xi$, abf, samples)
for $k = 1, 2, 3, \dots$ **do**
  $\lambda \leftarrow$ LAMBDA(ξ); $\delta\lambda \leftarrow$ LAMBDADERIVATIVE(ξ)
  $z_\xi \leftarrow z_\xi + \delta\lambda^2 \left[r_\xi^2 - \eta_\xi\right] / \gamma_\xi$
  $z_\theta \leftarrow z_\theta + \lambda^2 \left[\boldsymbol{r}_\theta^\top \boldsymbol{r}_{\theta_t} / \dim(\boldsymbol{r}_\theta) - \eta_\theta\right] / \gamma_\theta$
  $\mathscr{S} \leftarrow$ NEXTBATCH($\mathscr{D}, k$); $\delta A \leftarrow$ abf[ABFINDEXING(ξ)]
  $\tilde{U} \leftarrow$ MODELFORWARD($\boldsymbol{\theta}, \mathscr{S}$); $\tilde{\boldsymbol{f}} \leftarrow$ MODELBACKWARD($\boldsymbol{\theta}, \mathscr{S}$)
  $r_\xi \leftarrow r_\xi - \delta\lambda \left[\eta_\xi \tilde{U} + \mathcal{N}(0, 2c_\xi\eta_\xi)\right] - \delta\lambda^2 z_\xi r_\xi + \eta_\xi \delta A$
  $\boldsymbol{r}_\theta \leftarrow \boldsymbol{r}_\theta + \lambda \left[\eta_\theta \tilde{\boldsymbol{f}} + \mathcal{N}(0, 2c_\theta\eta_\theta \boldsymbol{I})\right] - \lambda^2 z_\theta \boldsymbol{r}_\theta$
  ABFUPDATE(abf, $\xi, \delta\lambda, \tilde{U}, k$)
  $\xi \leftarrow \xi + r_\xi$
  if ISINSIDEWELL(ξ) = **false then** ▹ ξ is restricted by the *well* of infinite height.
    $r_\xi \leftarrow -r_\xi$; $\xi \leftarrow \xi + r_\xi$ ▹ ξ bounces back when hitting the wall.
  $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \boldsymbol{r}_\theta$
  if $k = 0 \mod K_0$ and $\lambda = 0$ **then**
    APPEND(samples, $\boldsymbol{\theta}$) ▹ $\boldsymbol{\theta}$ is collected as a new sample in samples.
    $\boldsymbol{r}_\theta \sim \mathcal{N}(\boldsymbol{0}, \eta_\theta \boldsymbol{I})$ and $r_\xi \sim \mathcal{N}(0, \eta_\xi)$ ▹ $\boldsymbol{r}_\theta, r_\xi$ is *optionally* resampled.

function ABFUPDATE(abf, $\xi, \delta\lambda, \tilde{U}, k$)
  $j \leftarrow$ ABFINDEXING(ξ) ▹ ξ is mapped to the index j of the associated bin.
  $\text{abf}[j] \leftarrow [1 - 1/k]\,\text{abf}[j] + [1/k]\,\delta\lambda \cdot \tilde{U}$



Research Case II – TACT-HMC

Simulation: 2D mixture of Gaussian

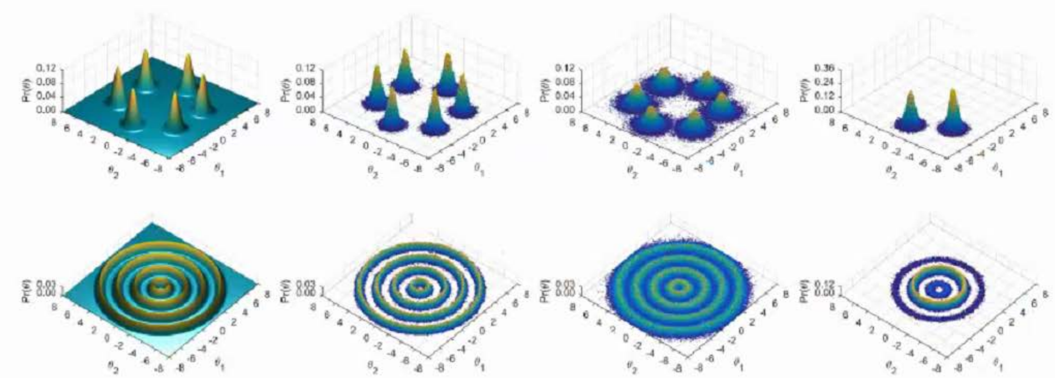


Figure 2: Experiments on sampling two $2d$ synthetic distributions. *Left*: The distributions to sample; *Mid-left*: Histogram of samples drawn by TACT-HMC; *Mid-right*: Histogram by the well-tempered sampler without thermostats; *Right*: Histogram by the thermostat-assisted sampler without tempering.

Deevlopment

Computer Vision: the field that designs algorithms and builds machines that can see

- The Advent of **Deep Learning** has disrupted ***Large-Scale Visual Recognition and Understanding***
- Deep Learning in CV: brain-inspired Deep Neural Networks (DNNs) trained on millions of labeled images ***achieve (super) human-level performance in object classification*** (e.g., Microsoft's PReLU-net in 2015 ImageNet 1000-class object classification)
- Deep Learning goes beyond object detection/recognition by ***unveiling context in the visual domain***
- ***Convolutional Neural Networks (CNN):*** main building-blocks that act as **multi-scale (unsupervised) representation learners** → no need for domain knowledge & handcrafted features

Example Images from the ImageNet ILSCVRC

- Transfer learning exploits structure among interrelated visual tasks: Reuse a pre-trained model on a new task
- Example : The tasks of object detection & segmentation are trained jointly → instance segmentation

"I think Transfer Learning is the key to General Intelligence",
Demis Hassabis, CEO, DeepMind

"Taskonomy: Disentangling Task Transfer Learning", Zamir et al., (best paper award CVPR 2018): ***explicitly model the structure of the visual tasks space***

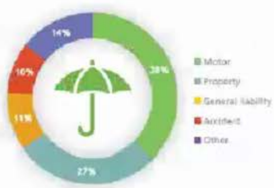
The number of data points needed for solving a set of 10 tasks can be reduced by 2/3 while keeping the performance the same !!!

Computer Vision Use Case I: Commercial Property (CP) Underwriting & Investment

Landscape: Property --- a significant portion of UW and Investment books --- faces multiple data issues

Non-Life and Property Insurance

Underwriting



- **European P&C** premiums totalled **€363bn** in 2016
- Property Premiums amounted to €99bn > **27% of the P&C total**.

Breakdown of Non-Life premiums in Europe 2016

Investments

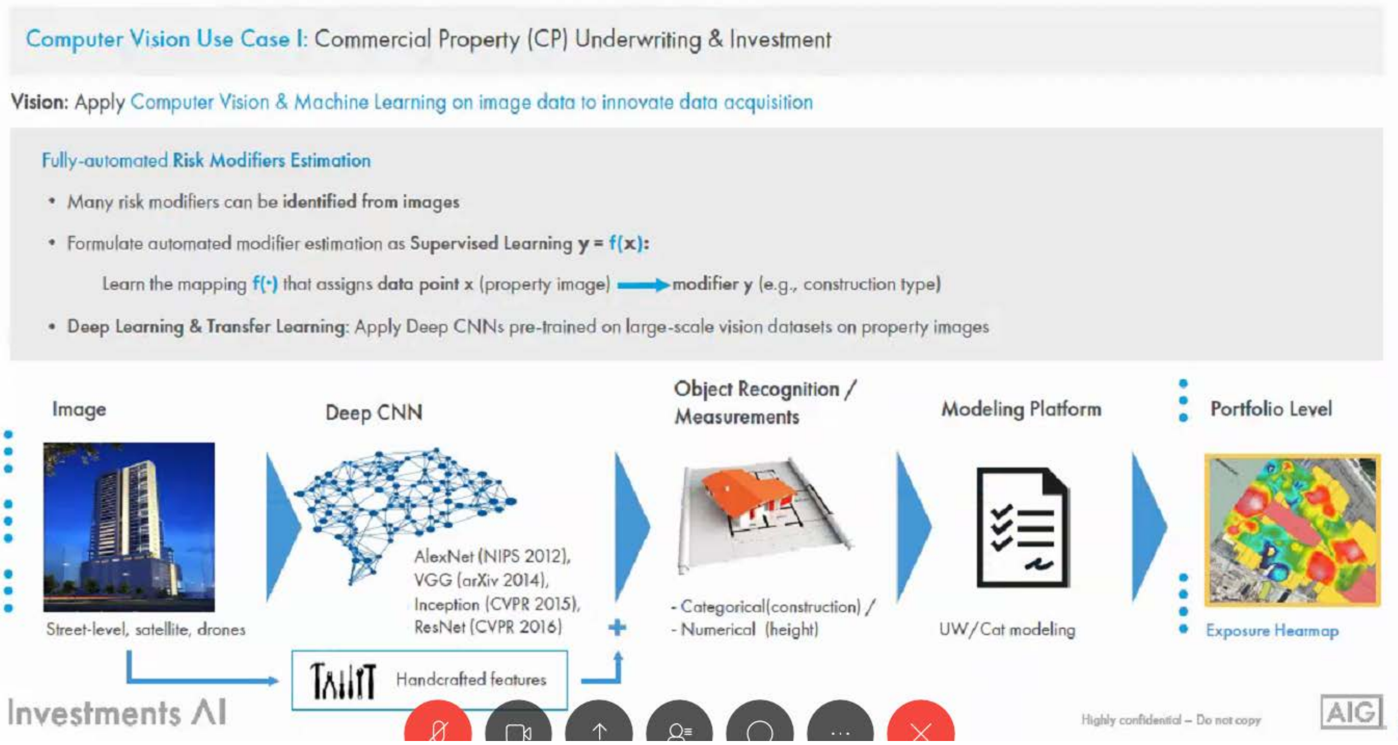
Insurers' investment portfolio in 2016 was equivalent to 61.9% of the EU's GDP

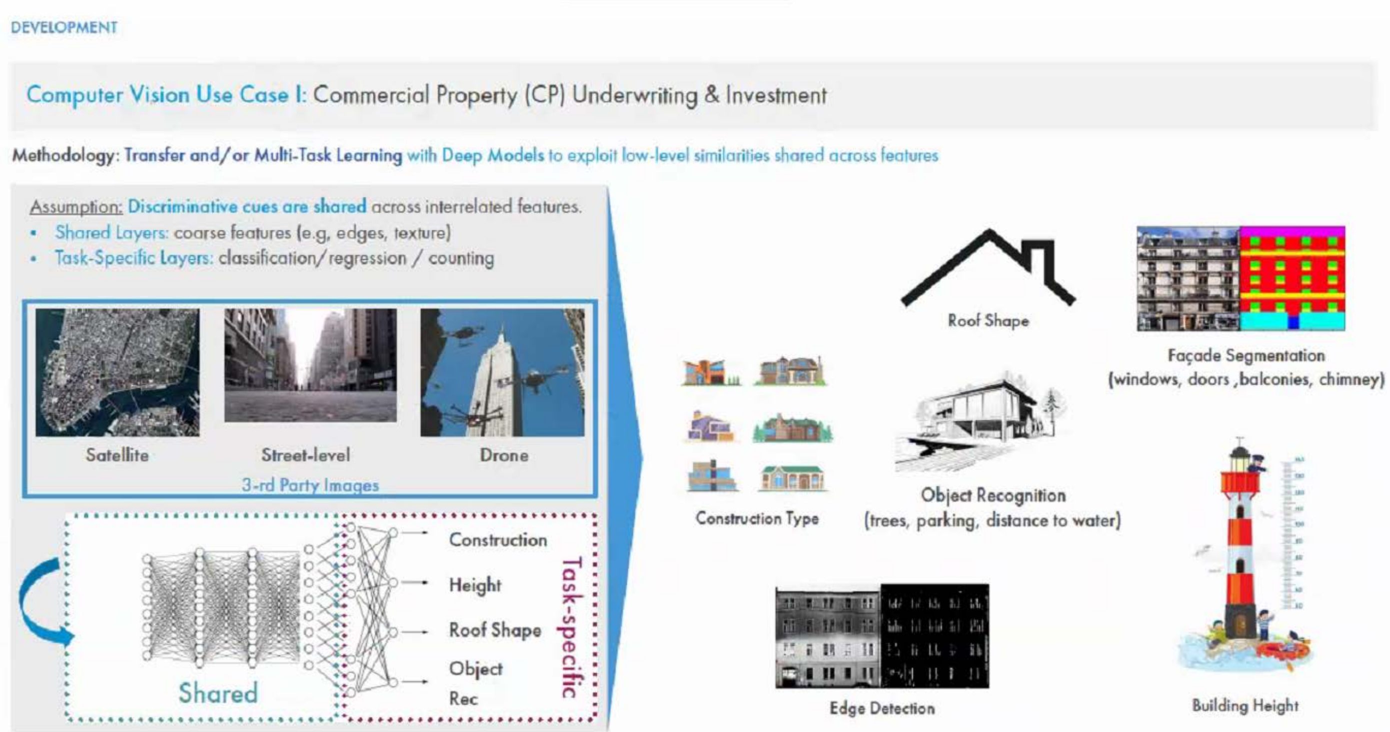
Commercial Property: attractive thanks to (a) higher yields, (b) long-term capital growth, (c) Better portfolio diversification, (d) flexible invest options (e.g., peer-to-peer, crowdfunding, funds and investment trusts)

Data Acquisition & Quality Issues in the Insurance Sector

1. **Property Underwriting** based on **primary risk modifiers** (e.g., construction type) **and secondary risk modifiers** of lower importance
2. **Unreliable & Expensive Data** collected from unreliable public records or pricy in-person inspections
3. **Low data quality** (incomplete, duplicate, inaccurate, outdated) due to the way business is transacted and lack of data sources

Detrimental impact on *response to market & portfolio analytics* → risk selection issues, cost/inefficiency of capital & RI purchasing, unknown exposure


Computer Vision Use Case I: Commercial Property (CP) Underwriting & Investment
Vision: Apply Computer Vision & Machine Learning on image data to innovate data acquisition
Fully-automated Risk Modifiers Estimation
• Many risk modifiers can be identified from images
• Formulate automated modifier estimation as Supervised Learning y = f(x):
Learn the mapping f(•) that assigns data point x (property image) → modifier y (e.g., construction type)
• Deep Learning & Transfer Learning: Apply Deep CNNs pre-trained on large-scale vision datasets on property images
Image
Deep CNN
Object Recognition / Measurements
Modeling Platform
Portfolio Level
Street-level, satellite, drones
AlexNet (NIPS 2012), VGG (arXiv 2014), Inception (CVPR 2015), ResNet (CVPR 2016)
- Categorical(construction) /
- Numerical (height)
UW/Cat modeling
Exposure Heatmap
Handcrafted features
Investments AI
Highly confidential – Do not copy
AIG


DEVELOPMENT
Computer Vision Use Case I: Commercial Property (CP) Underwriting & Investment
Methodology: Transfer and/or Multi-Task Learning with Deep Models to exploit low-level similarities shared across features
Assumption: Discriminative cues are shared across interrelated features.
• Shared Layers: coarse features (e.g, edges, texture)
• Task-Specific Layers: classification/ regression / counting
Satellite
Street-level
Drone
3-rd Party Images
Shared
Construction
Height
Roof Shape
Object Rec
Task-specific
Roof Shape
Façade Segmentation (windows, doors ,balconies, chimney)
Construction Type
Object Recognition (trees, parking, distance to water)
Edge Detection
Building Height

Computer Vision Use Case II: Auto Damage Detection

DEVELOPMENT

Computer Vision Use Case II: Auto Damage Detection

CRITICAL PAINPOINTS FOR THE CLAIM INDUSTRY

High Processing Cost: costs due to ***labour-intensive appraisal procedure*** & **Elongated Claim Processing**: customer dissatisfaction → ***costly turnover***

I. Significant Processing Cost*

- North America, 2016: **26M auto claims with $22B total claim payout**
- **Three major processing costs** account for 32% of the total claim payout (~$7.0B):
  1. **Car Rentals** ($3,7B): 40% of claims, Avg. 12 days at $30/day
  2. **Vehicle Storage** ($1.5B): 4.4M total loss claims spent an Avg. of $344/claim in storage fees (Avg. 17 days)
  3. **Adjusters** ($1.8B):
     - Repairable claims: **30% Independent Adjusters** at 150/claim, **24% Insurers' Staffed Adjusters** (11k employed in US at $65,000/year), **4% Independent Photo Appraisers** at $150/claim
     - Frequent supplemental appraisal (>50%) costing $125/claim

II. Elongated Claim Processing

- **5-15 days for Repairable Claims , 30 days for Total Loss Claims**
- **88%** of displeased claimants **avoided the "definitely will renew"** response and **93% avoided the "definitely will recommend"** their current insurer (J.D. Power 2016 U.S. Auto Claims Satisfaction Study)
- **7x– 9x higher cost to attract a new customer than to retain one**
- 2014: traditional auto insurance companies spent an avg. of **$792/ policy on new customer acquisition**

Computer Vision Use Case II: Auto Damage Detection

Automatic Car Damage Detection system

Summary

Automatic Car Damage Detection: end-to-end *Computer Vision* and *Machine Learning* solution that prepares an assessment of damages to a vehicle when provided with images of the exterior of the vehicle along with the vehicle's make, model and year.

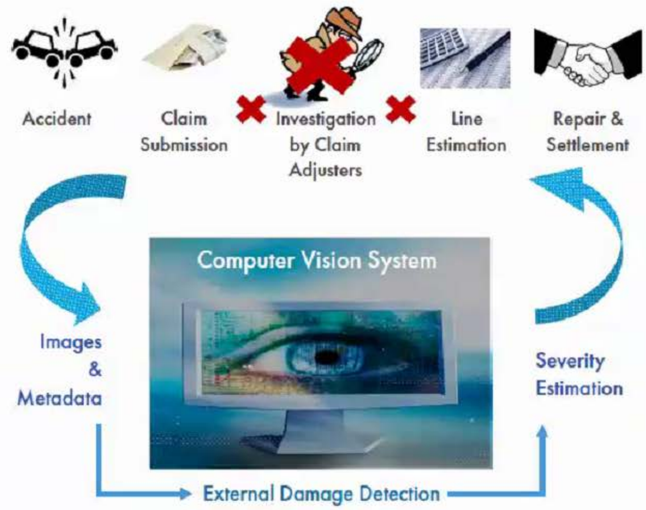
Product

A. **Convolutional Neural Networks (CNN)** ***learn to assess exterior***

B. Coupled with a database of parts and labour cost, the list of damaged parts thus predicted can be trained to ***predict the parts and labour cost.***

Use Cases

- Automated Repair Cost Estimation & Total Loss Prediction
- Repair shop Early Notification
- Reduction in Storage Time:
- Improved Customer Satisfaction
- Reduction of Fraud



Yuanyuan Liu, D.Phil. (Oxon)

Director, Statistical Machine Learning
Investments AI | Investment
+44-2076-516247
yuanyuan.liu@aig.com

Thanks