Swiss Re Institute

Addressing the failure of enterprise machine intelligence (MI) with 4 AIs we should discuss more

Dr. Jeffrey Bohn, Chief Research & Innovation Officer

January, 2020

Swiss Re Institute

Why the Swiss Re Institute?

Superior research driving better decisions

- Support internal decision making processes
- Commercial leverage through supporting clients and stakeholders

institute.swissre.com

How we define "artificial" and "intelligence" will influence research, development, and deployment of machine intelligence (MI)

- **Artificial**: Human-made, contrived, not natural, not real (*maybe pseudo- is a better adjective as it explicitly means "to resemble or imitate"*)

- **Intelligence**: Learn & **apply knowledge**/skills, sol reason/ plan/ **adapt**/ respond/ think abstractly (*sh from Latin intelligere that means "to understand"*)

- **How we define machine intelligence materially influences the tools we create**

- **Augmented intelligence** & **intelligent automation** portend more productive futures

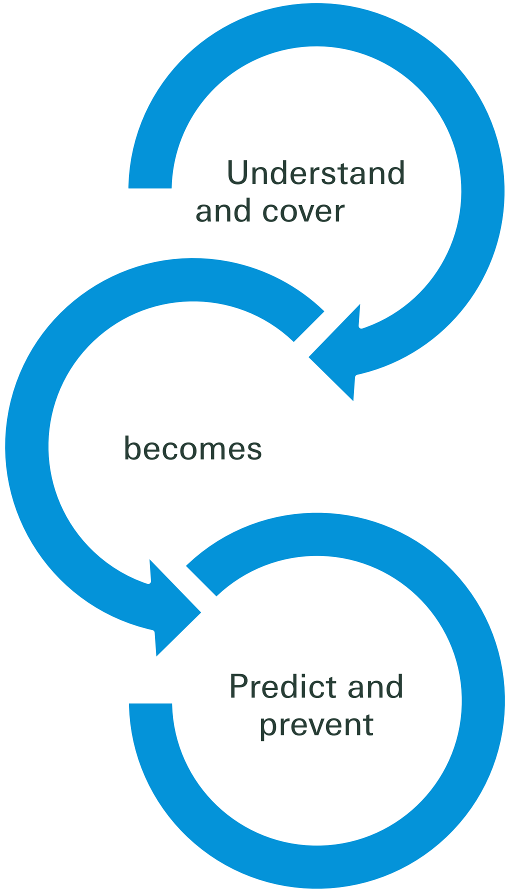
- **Machine** strengths: *Literal*
  - Computation
  - Consistent
  - Ubiquitous
  - Detailed
  - Networkable (at scale)

- **Human** strengths: *Heuristic*
  - Adaptive
  - Autonomous
  - Tacit/subtle understanding
  - Imaginative
  - Emotional/social understanding
  - Robust

The future will likely still be Human Intelligence augmented by some kind of Machine Intelligence

The promise of MI in insurance

Shifts in the insurance industry



Machine learning (ML) & artificial intelligence (AI) successes: Anecdotal evidence

Ant Financial – Claims Assessment

"...a user can take a photo of their vehicle damage using their smartphone camera or upload photos of the damage into the software...the AI "assessed" the damages and handled the claims in six seconds. The human claims adjusters apparently took six minutes and 48 seconds to reach their conclusions."

Tokio Marine – Document Processing

"Tokio Marine has an AI-assisted claim document recognition system that helps to handle handwritten claims notice documents using a cloud-based AI optical character recognition (OCR) service. It reduces 50 percent of the document input load as well as complies with privacy regulations...The results: over 90 percent recognition rate, 50 percent reduction in input time, 80 percent reduction in human error..."

Progressive - Optimized Insurance Rates

"Progressive in turn receives...reportedly over 14 billion miles worth [of driving data]...that its machine learning software can use to extract the patterns that exist between a driver's demographics and their driving habits and then offer a rate to customers that is tailored specifically to them..."

1. https://emerj.com/ai-sector-overviews/ai-auto-insurance-current-applications/
2. https://www.accenture.com/t20180822t093440z__w__/us-en/_acnmedia/pdf-84/accenture-machine-leaning-insurance.pdf
3. https://emerj.com/partner-content/connected-insurance-and-ai-the-possibilities-of-iot-data/

The failure of enterprise MI

Machine learning (ML) & artificial intelligence (AI) failures: Anecdotal evidence

- MD Anderson Cancer Center and IBM Watson[1]
  - "…it was suggested that the program isn't usable in most cases."[2]
- Amazon AI Recruiter
  - "…had to be scrapped after showing a distinct bias against women."[2,3]
- Yandex's Alice Chatbot
  - "…mentioned pro-Stalin views; support for wife-beating, child abuse and suicide."[4]
- Amazon Echo
  - A television news report about an Echo accidentally ordering a dollhouse triggered many other active Echos listening to the television to automatically order dollhouses.[5]
- Microsoft Cortana
  - Failed to recognize non-American accents.[6]

1. https://towardsdatascience.com/ai-fails-and-what-they-teach-us-about-emerging-technology-e6978c8c4922
2. https://www.itproportal.com/features/ai-fails-why-ai-still-isnt-ready-to-take-your-job/
3. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G
4. https://techcrunch.com/2017/10/24/another-ai-chatbot-shown-spouting-offensive-views/
5. https://usa.kaspersky.com/blog/voice-recognition-threats/10855/
6. https://youtu.be/DDqrfCmIPxI

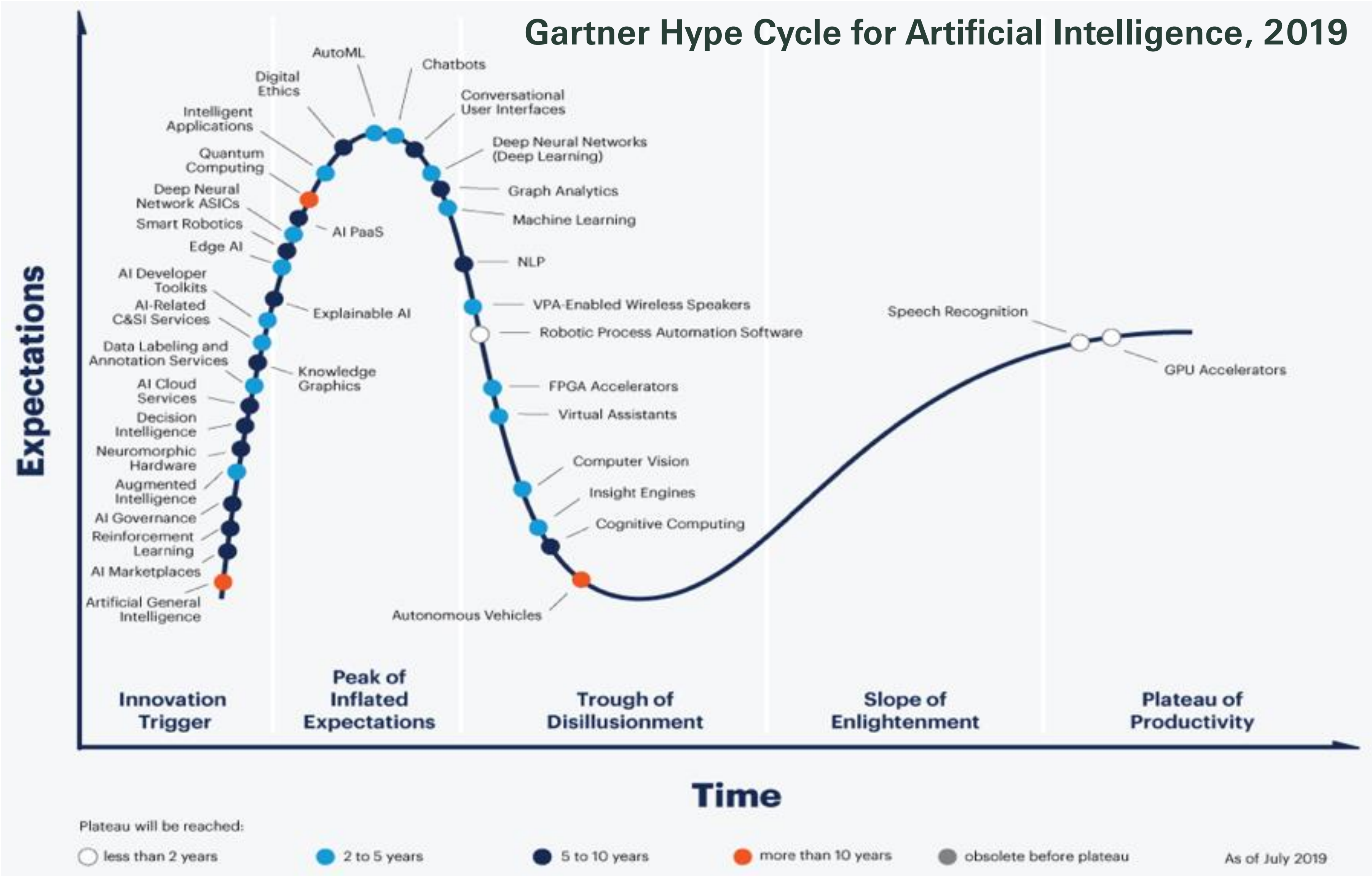
The reality:

88% of AI projects in a 2017 survey did not progress beyond the experimental stage.[1]

1. MLOps: Machine Learning Operationalization (ActiveState) https://www.activestate.com/wp-content/uploads/2018/10/webinar-slides-mlops.pdf

Gartner estimates over 85% of AI projects fail.

Source: https://www.techrepublic.com/article/85-of-big-data-projects-fail-but-your-developers-can-help-yours-succeed/



Swiss Re Institute

Why do so many MI projects fail?

- **Misconceptions that AI works in the same way the human brain does: Algorithms are typically hyper-narrowly focused**
- **Poor decision-maker or customer adoption: Output not well integrated into existing decision processes**
- **AI/ML projects differ from most other IT projects in that they often require domain expertise on the project team: Most projects still run as if the approach is just one more tool to be implemented**
- **Algorithms tend to be opaque and post-mortem assessments can be difficult when decisions are wrong: Often, it can be hard to decide where a problem developed– data curation? training/calibration? algorithm implementation? process engineering?**
- **AI/ML is not the right tool for every forecasting task: Model complexity should fit the use case– Occam's razor in its original form i.e., no more structure/complexity than what is necessary**
- **Hard to find right talent: Demand exceeds supply for software architects & quality AI/ML programmers**

Four AIs we should discuss & use more to address enterprise MI failure...

- **Augmented** Intelligence: Improve productivity now

- Intelligent **Automation:** Keep humans in the loop to build optimal systems

- **Assessed** Intelligence: Monitor & evaluate algorithms/models/systems on a regular basis

- **Adaptive** Intelligence: Build systems that adapt (path forward likely to use rigorous causal inference)

Augmented Intelligence

Synthesize unstructured data and combine with structured data to improve risk assessment

- Quantitative models– particularly, market-based-analytics outperform fundamental analysts 80 to 90% of the time

- Assessing performance of human analysts that outperform MI can lead to improved hybrid approaches

- ***Key solutions often arise from augmenting human analysis by…***
  - ***Synthesizing unstructured data to reduce noise-to-signal ratio***
  - ***Incorporating non-financial-statement-related unstructured data (e.g., ESG)***
  - ***Filtering structured data to focus analysts on exceptional & subtle characteristics where humans excel***
  - ***Incorporating near-real-time performance assessment & calibration***
  - ***Comparing, contrasting, curating, and/or combining standard analytics with machine-learning-enabled approaches***

96% of respondents in a recent survey indicated “a lack of training data, technology, and skills has impeded their ability to train their ML algorithms and attain the confidence their model must provide.”

Source: https://dataconomy.com/2019/07/why-96-of-enterprises-face-ai-training-data-issues/

Tools for better augmenting human intelligence

https://jupyter.org/

JupyterLab is a web-based interactive development environment for Jupyter notebooks, code, and data.

https://github.com/jupyterhub/binderhub

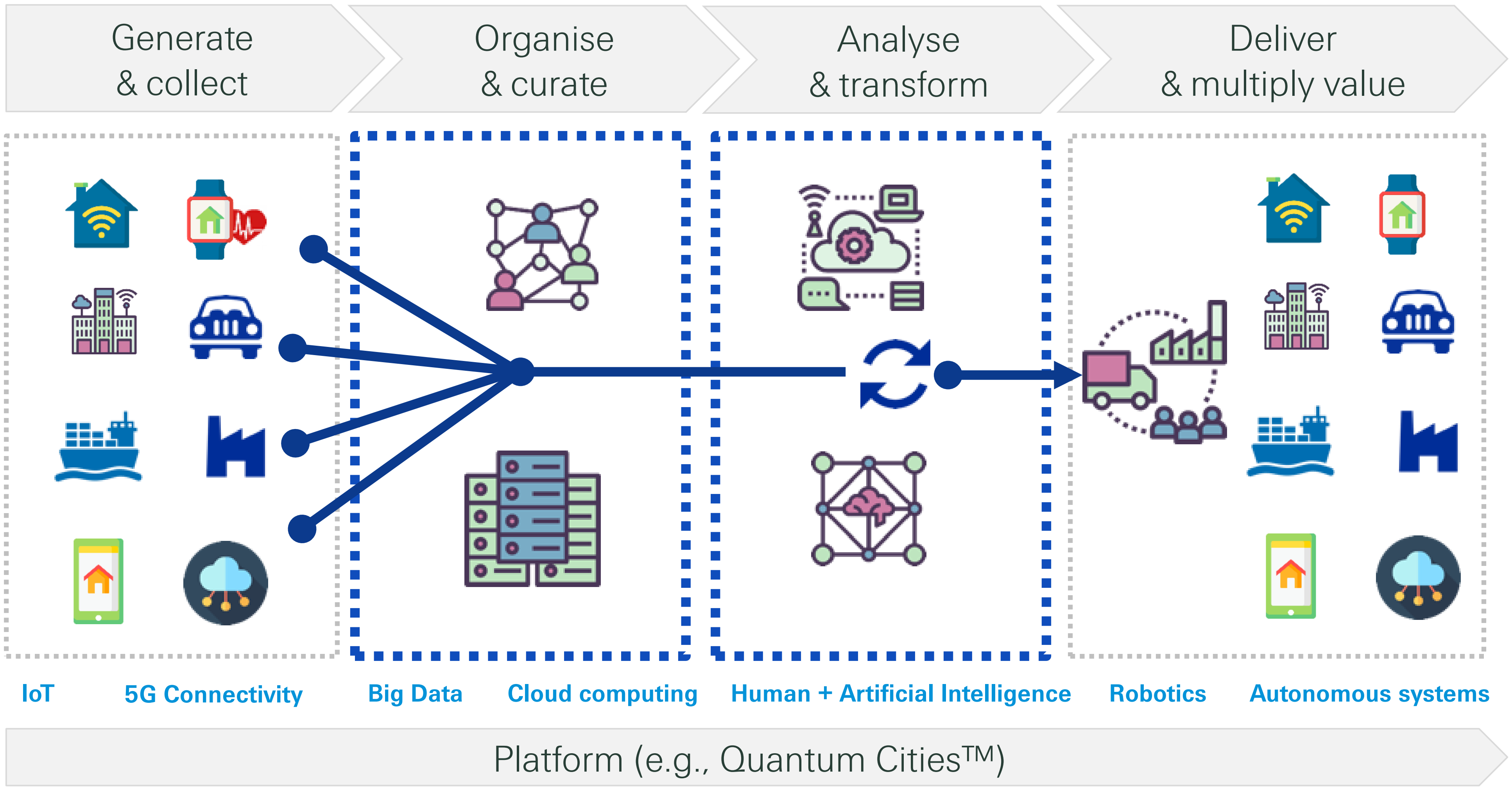
BinderHub allows you to *build* and *register* a Docker image from a Git repository, then *connect* with JupyterHub, allowing you to create a public IP address that allows users to interact with the code and environment within a live JupyterHub instance.

https://mlflow.org/

MLflow is an open source platform to manage the ML lifecycle, including experimentation, reproducibility and deployment.

Intelligent Automation

Data is the oil for machine intelligence: Crude multiplies in value after refinement



Swiss Re Institute

Improving data ingestion & curation

- Centralize data ingestion & curation
- Shift paradigm for data from a stock asset to a flow asset– connect to as many data pipes as possible
- Collect & deploy filtering rules for better curation
  - Missing data approach
  - Outlier handling
  - Overwriting policies
  - Test data against rules for *range possibility, range plausibility, relational plausibility, & pattern plausibility*

Use machine intelligence to intelligently automate data ingestion & curation

Many traditional enterprises have capable **DevOps** resources & processes; however, **MlOps** for machine intelligence are relatively new.

Automated underwriting often leads to simple errors when humans are taken out of the "loop"

Assessed Intelligence

Brave New World author Aldous Huxley at U.C. Berkeley in 1962:

Our business is to be aware…and then use our imagination to see what might happen, how this might be abused and…see that…scientific and technological advances be used for the benefit of human beings and not for their degradation.

Machine intelligence assessment is not just about performance...
Criteria to evaluate a new end-to-end process should include the following:

- Model performance at multiple, relevant time horizons
- Re-engineered end-to-end process robustness & related diagnosability
- Acceptable processing time for re-engineered end-to-end process
- Necessary data reliability & cost
- Availability & ease of hiring relevant talent
- Algorithmic risk & malpractice (likelihood) in terms of system vulnerability– cyber risk & adversarial ML risk--, system fragility, bias, & regulatory compliance
- Productive workflow transformation assessed as greater (in terms of reduced cost, increased revenue, or options for new opportunities) than the cost of organizational integration (both direct development & running costs and indirect organizational & opportunity costs)

Note that slightly worse performance may be acceptable when a new end-to-end process is evaluated across all these dimensions

Adaptive Intelligence

Hybrid modeling

- Post-disaster resilience modeling
  - Physical model
  - Reinforcement learning

- Time-series prediction of financial variables
  - Reinforcement learning
  - Data augmentation
  - Behavior mimicking

Shifting to a causal-inference paradigm creates more adaptive intelligence

- A/B testing or randomized controlled trials (RCTs) are the first choice— not always possible
- Causal inference moves along the following causal hierarchy:
  - Association
  - Intervention
  - Counterfactuals
- Recommendations[1]
  - Encode causal assumptions/hypotheses using graphical models
  - Control confounding (*Pearl´s Do-calculus*)
  - Algorithmize counterfactuals (shift from “effects of causes” to “causes of effects”)
  - Analyze intermediate mechanisms and assess both direct & indirect effects
  - Adapt & validate causal outcomes from one environment to another
  - Search systematically & prune compatible models i.e., discover causal compatibility of hypothesized graph & data

Causal inference may be the new frontier as we migrate from association-based analysis, only

1. Pearl, Judea, November 2018, “The seven tools of causal inference with reflections on machine learning,” Technical Report R-481, *Association for Computing Machinery*

Adapting MI to address extreme-downside scenarios

Scenario analyses & simulation are under-used to augment decision making

Extreme-downside, scenario categories are not created equal

- Black swans (Nassim Taleb): Unknowable given current information & virtually impossible to predict
- Gray rhinos (Michele Wucker): Highly probable & straightforwardly predictable, but neglected
- Perfect storms (Operational research): Low probability & not straightforwardly predictable given that the outcome results from interaction of infrequent events, but can be identified via scenario analysis

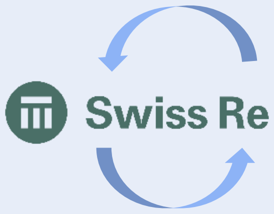
Appendix

Examples where digitizing trends changes insurance

- Forward-looking modeling of risk pools
- Incorporating unstructured data into business and capital steering
- Tracking natural catastrophe damage in real time
- Assessing damage
- Automated underwriting
- Improving customer targeting
- Parametric insurance contract implementation
- Intelligent automation & robotic process automation (RPA) for underwriting and claims processing
- Chatbots for customer support
- Natural language processing applied to contract review

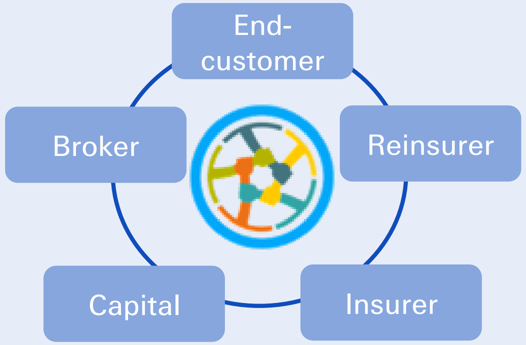
Changes arising from distributed ledger technology (DLT) & machine intelligence (MI)

Internal improvements



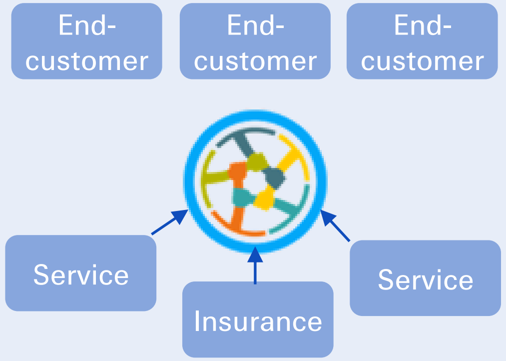
Enhancing **internal business process efficiencies** by hosting shared services on internal DLT platforms

Enhancement of insurance value-chain



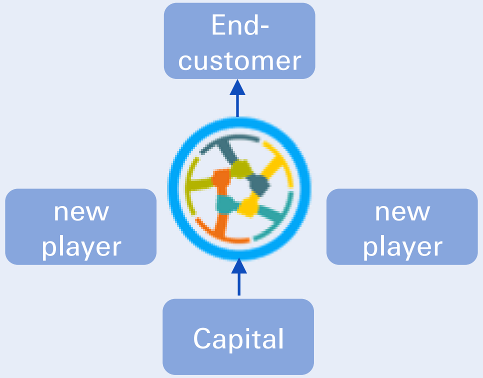
Organizing members of **existing insurance value-chain** into a DLT cooperative for driving data standardization, aggregated data access and shared processes

End-customer driven initiatives



End-customers in **non-insurance verticals** organizing DLT cooperatives for aggregation of shared services including insurance

New value-chains



New players creating a **DLT-based alternative value-chain** or market for distribution of re/insurance services

Incremental efficiency gains ⟷ *Business model disruption*

Machine learning's successes and failures

Successes

- Facial recognition
- Radiology
- Autonomous driving
- Advertising
- Low-impact recommender systems

Failures

- Economic & financial time series prediction
- High-impact recommender systems
- Customer service systems
- End-to-end transformation of labor-intensive workflow in governments & companies
- Personal digital assistants
- Robo-advising
- Health interventions

Contrasting DevOps and MlOps

	DevOps	MlOps
Automate tasks where possible.	✓	✓
Provision, maintain and monitor diverse set of system/cloud resources.	✓	✓
Strives to ensure scalability.	✓	✓
CI/CD for applications is more common.	✓	
Heavy emphasis on data preprocessing and *data* version control.		✓
Highly complex (and ever-changing) package/library dependencies.		✓
More common to rely upon container-based deployments.		✓
Often highly sophisticated, large scale, distributed data stores.		✓
Focus on model train/test accuracy before production deployment.		✓
Wider array of training and certifications available[1].	✓	

1. As of October 28, 2019 a Google search for "DevOps certification" returns 49m~ results while a search for "MLOps certification" returns 8,730 results. Similarly, a search for the same terms on monster.com yields 2,535 and 2 results, respectively.

Legal notice