

Anti-data-science

the Taleb's barbell in the AI world

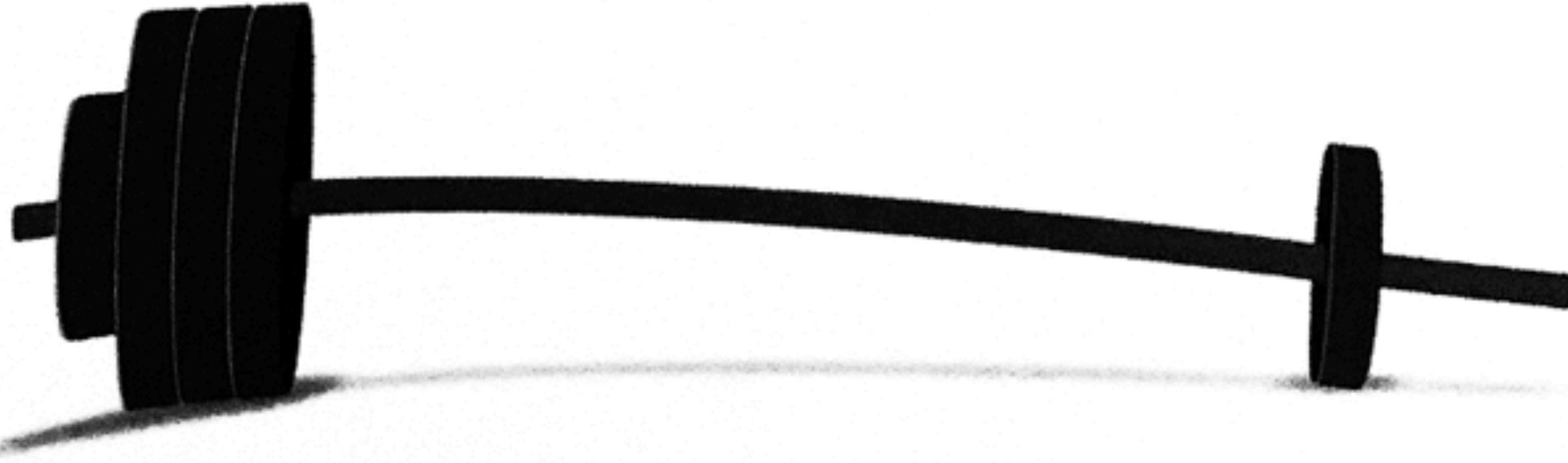
Why I am here?

- ~~Bla-bla-bla about company I've founded, universities I've taught in, systems I've built, articles published, etc~~
- What novel and contrarian I can bring to the table:
 - how to build a company with culture of independent experts who are engaged in projects with other institutions
 - **why to build data science and AI solutions where there is no data and no AI**

The Barbell Strategy (Taleb, 2007)

Low risk
(the known)

High risk
(the unknown)



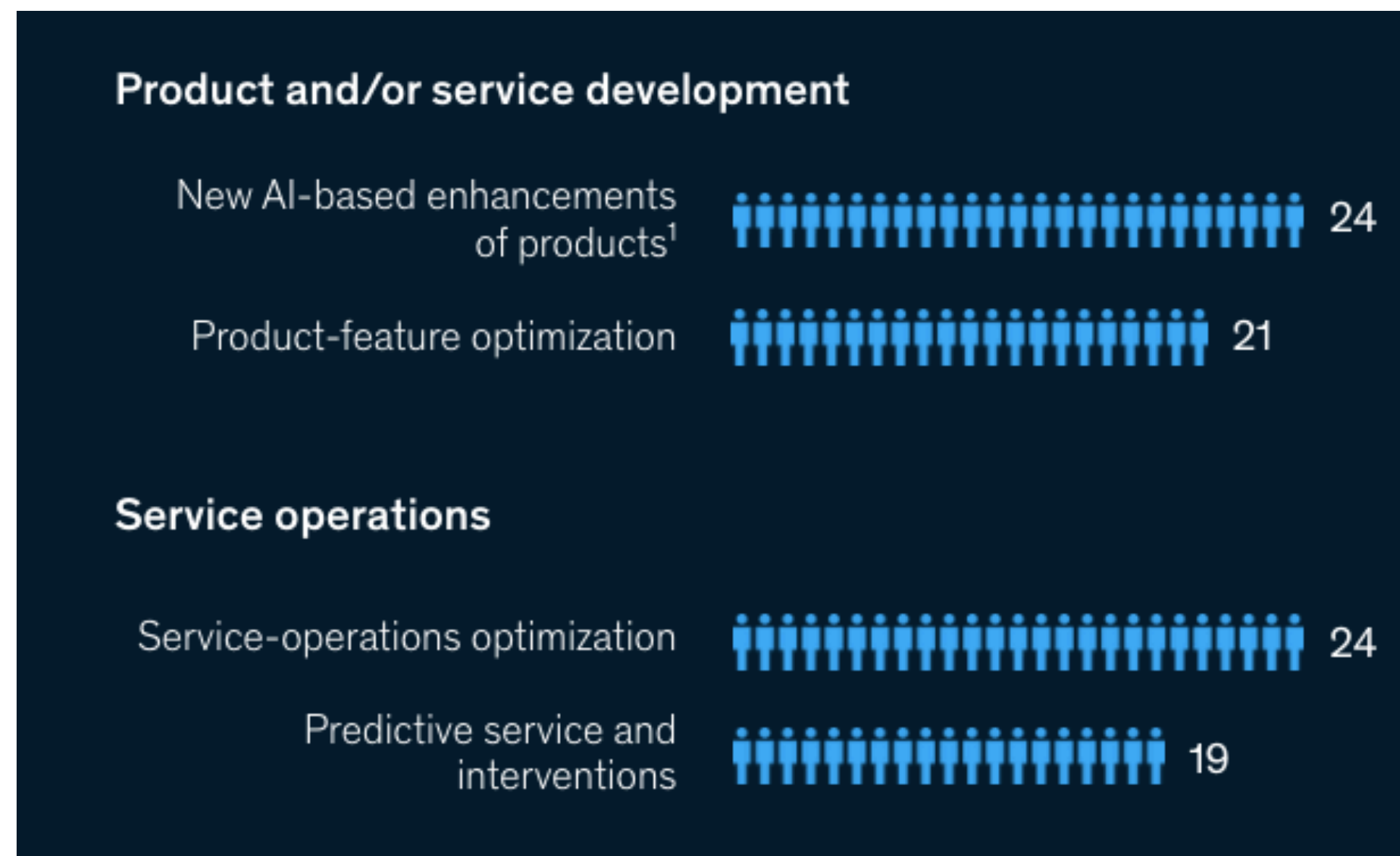
Extremely conservative
investments
(90%)

Extremely aggressive
investments
(10%)

**Extremely conservative
investments in AI**

Business of AI

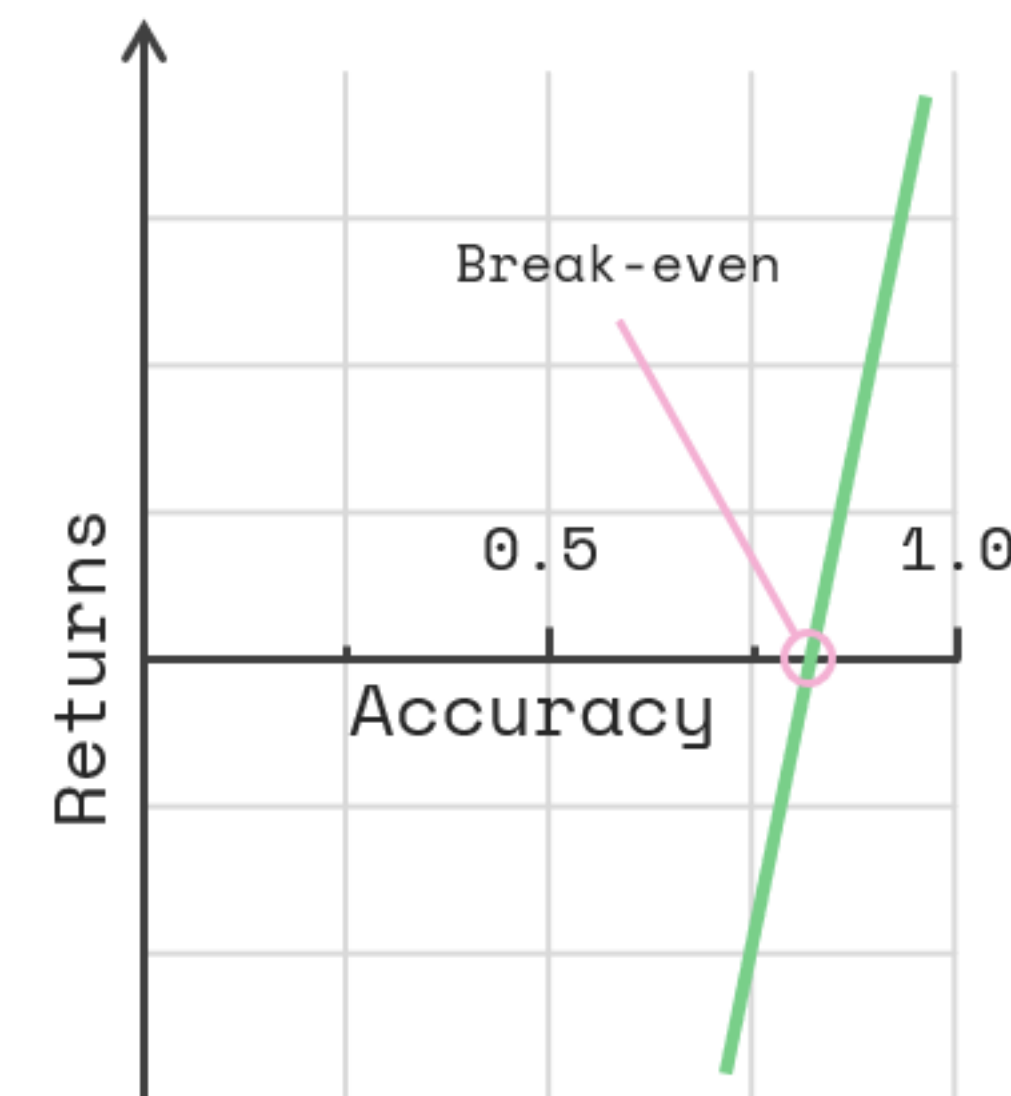
Conservative big data



<https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2020>

$$\text{returns} = \text{value} - (1 - \text{accuracy}) * \text{cost of a mistake}$$

$$\text{accuracy} = 1 - \frac{\text{value}}{\text{cost of a mistake}}$$



<https://towardsdatascience.com/return-on-investment-for-machine-learning-1a0c431509e>

Business of AI

Conservative big data

Category	Conservative big data	Aggressive zero-data
The “world”	Bits and bytes	
The process	Routine, often intuitive tasks (System 1)	
Capitalization / moats	Massive data	
Unit economics	Linear / logarithmic improvements	

Technology of AI

Conservative big data

<i>Dataset</i>	<i>Quantity (tokens)</i>	<i>Weight in training mix</i>	<i>Epochs elapsed when training for 300B tokens</i>
<i>Common Crawl (filtered)</i>	410 billion	60%	0.44
<i>WebText2</i>	19 billion	22%	2.9
<i>Books1</i>	12 billion	8%	1.9
<i>Books2</i>	55 billion	8%	0.43
<i>Wikipedia</i>	3 billion	3%	3.4

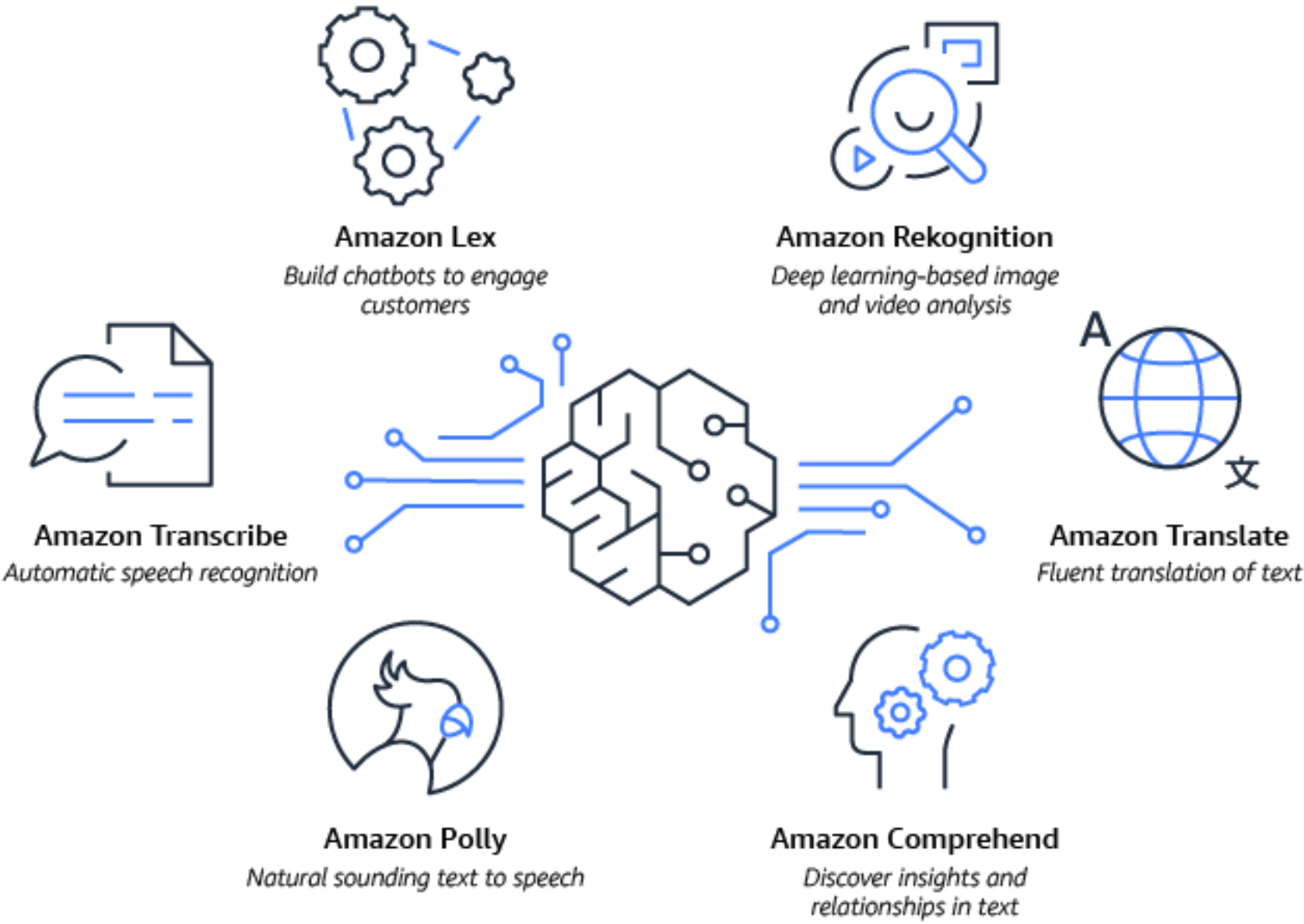
<https://in.springboard.com/blog/openai-gpt-3/>

GPT Playground

File Edit View Insert Format Data Tools Add-ons Help Last edit wa

100% \$ % .0 .00 123 Default (Ari... 10

	A	B	C	D
1	State Name	Population		
2	Illinois	12.67M		
3	California	39.51M		
4	Ohio	11.69M		
5	Michigan			
6				



Technology of AI

Conservative big data

Category	Conservative big data	Aggressive zero-data
Datasets size	Millions, billions samples	
Targets and metrics	Single optimization criteria	
Mathematical problem	Pattern recognition	
Algorithms	Popular, commoditized, ready-to-use	

People of AI

Conservative big data

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21st century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- ☆ Machine learning
- ☆ Statistical modeling
- ☆ Experiment design
- ☆ Bayesian inference
- ☆ Supervised learning: decision trees, random forests, logistic regression
- ☆ Unsupervised learning: clustering, dimensionality reduction
- ☆ Optimization: gradient descent and variants

PROGRAMMING & DATABASE


- ☆ Computer science fundamentals
- ☆ Scripting language e.g. Python
- ☆ Statistical computing package e.g. R
- ☆ Databases SQL and NoSQL
- ☆ Relational algebra
- ☆ Parallel databases and parallel query processing
- ☆ MapReduce concepts
- ☆ Hadoop and Hive/Pig
- ☆ Custom reducers
- ☆ Experience with xaaS like AWS


DOMAIN KNOWLEDGE & SOFT SKILLS

- ☆ Passionate about the business
- ☆ Curious about data
- ☆ Influence without authority
- ☆ Hacker mindset
- ☆ Problem solver
- ☆ Strategic, proactive, creative, innovative and collaborative


COMMUNICATION & VISUALIZATION

- ☆ Able to engage with senior management
- ☆ Story telling skills
- ☆ Translate data-driven insights into decisions and actions
- ☆ Visual art design
- ☆ R packages like ggplot or lattice
- ☆ Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau






Max Lytvyn





Alex Shevchenko




Dmytro Lider

<https://www.grammarly.com/about>







Oleg Rogynskyy
Chief Executive Officer & Founder



Justin Shriber
Chief Marketing Officer



Art Harding
Chief Operating Officer



Thomas Wyatt
Chief Product and Strategy Officer

<https://people.ai/about-us/>

<https://www.kdnuggets.com/2020/01/wanna-be-data-scientist.html>

People of AI

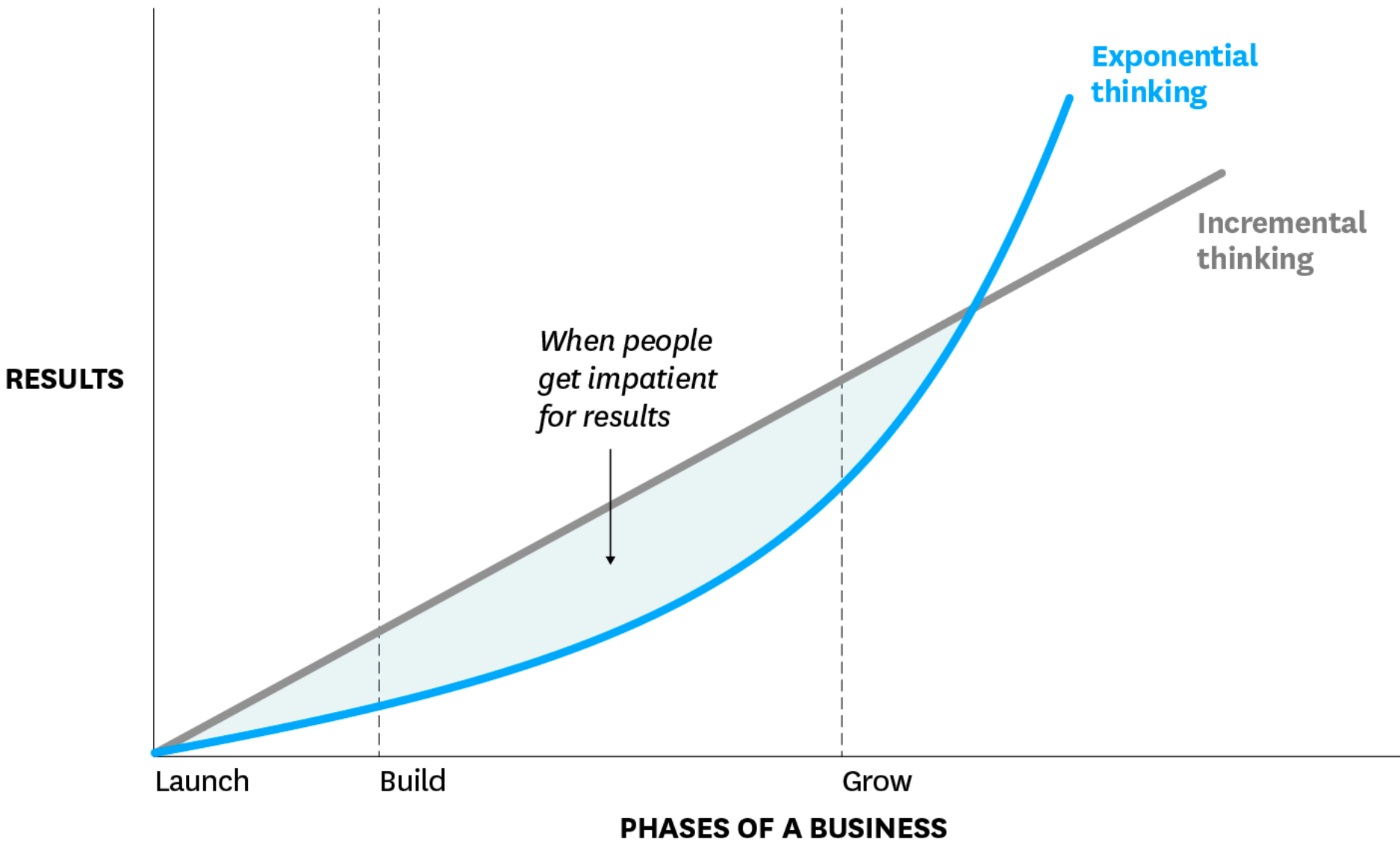
Conservative big data

Category	Conservative big data	Aggressive zero-data
Stakeholders	Business, product, marketing	
Developers	Engineers-integrators	
Customers	“Regular users”	

**Extremely aggressive
investments in AI**

Incremental vs. Exponential Thinking When Growing a Business

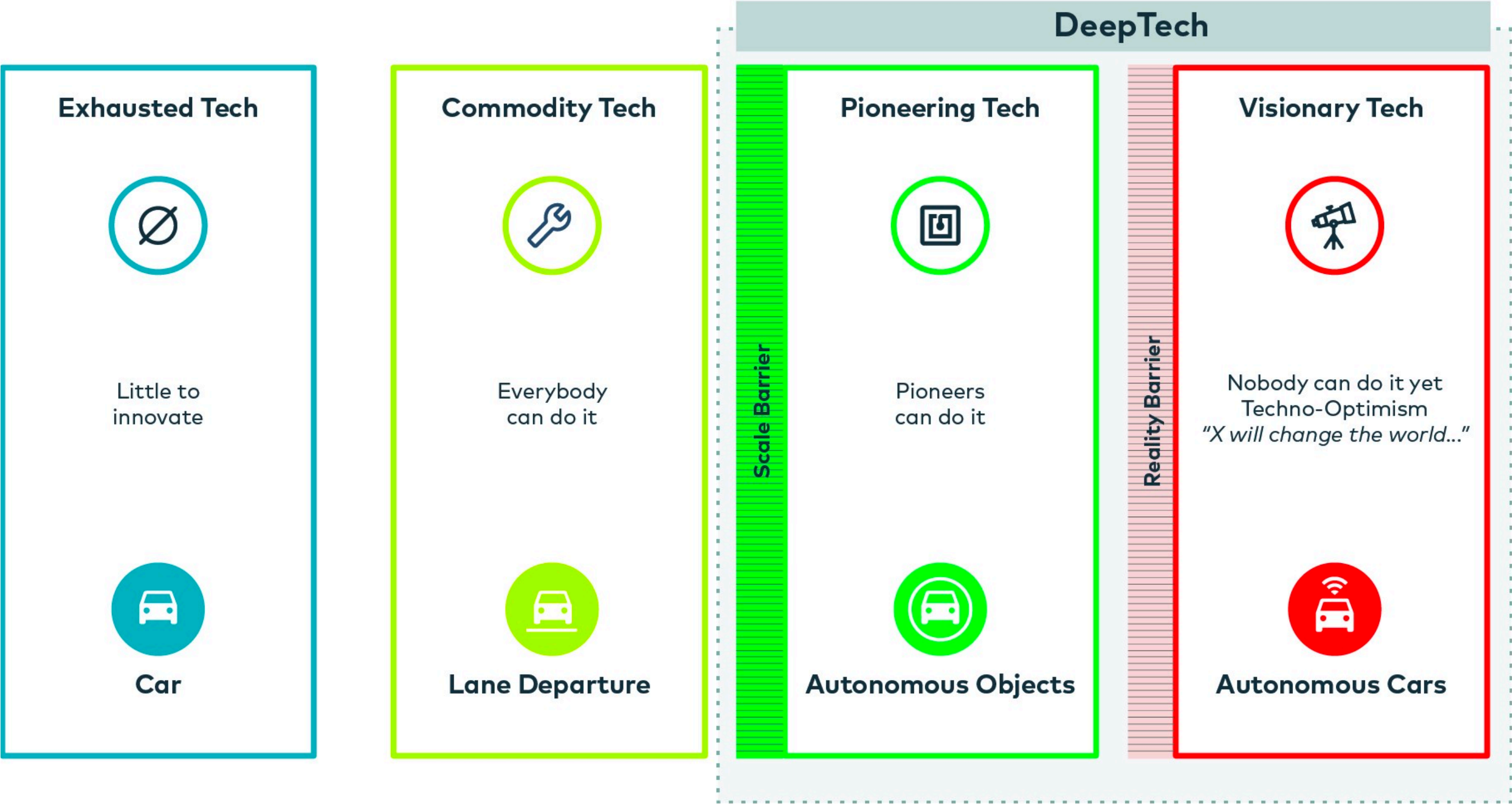
Incremental thinking delivers immediate and steady results, while exponential thinking generates results that accelerate over time. The wrong expectations can lead teams to quit the exponential path too soon.



SOURCE MARK BONCHEK

© HBR.ORG

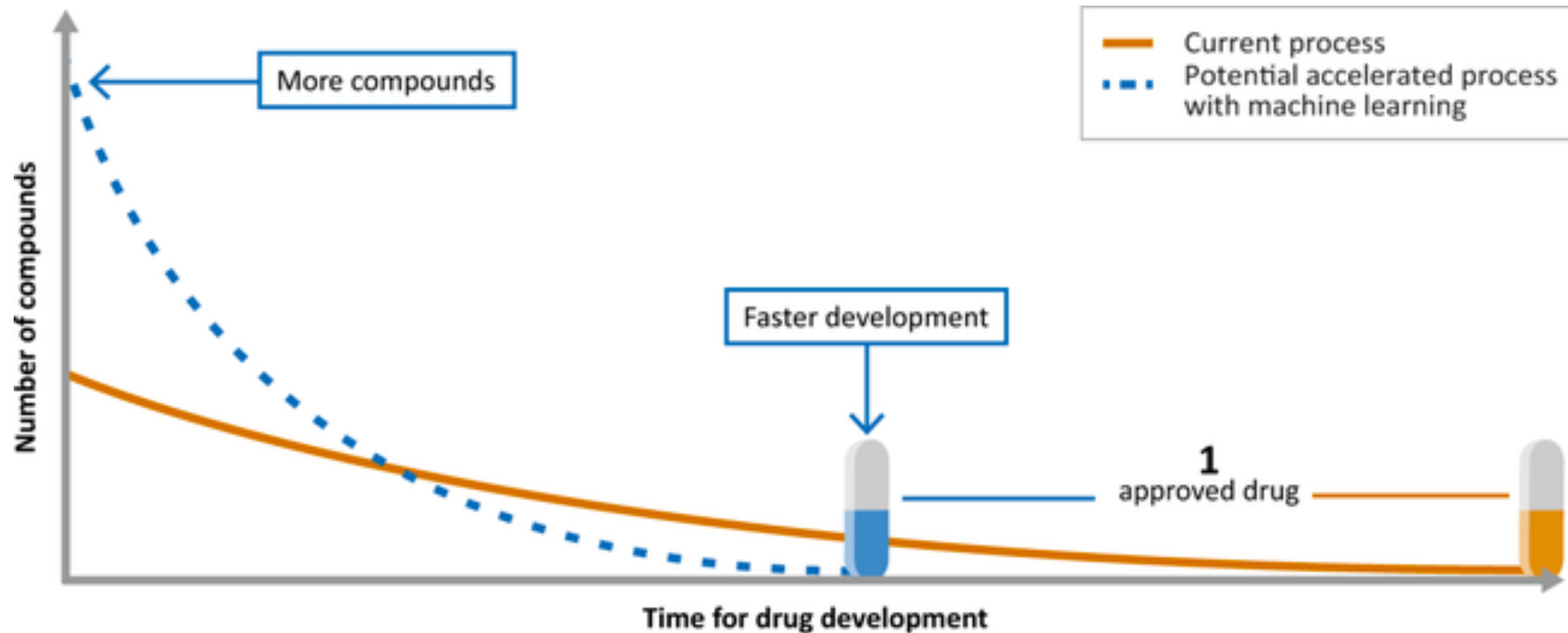
<https://hbr.org/2016/07/how-to-create-an-exponential-mindset>



<https://medium.com/bcg-digital-ventures/the-right-time-for-deep-tech-dcb317fc3636>

Business of AI

Aggressive zero-data



Source: GAO. | GAO-20-215SP

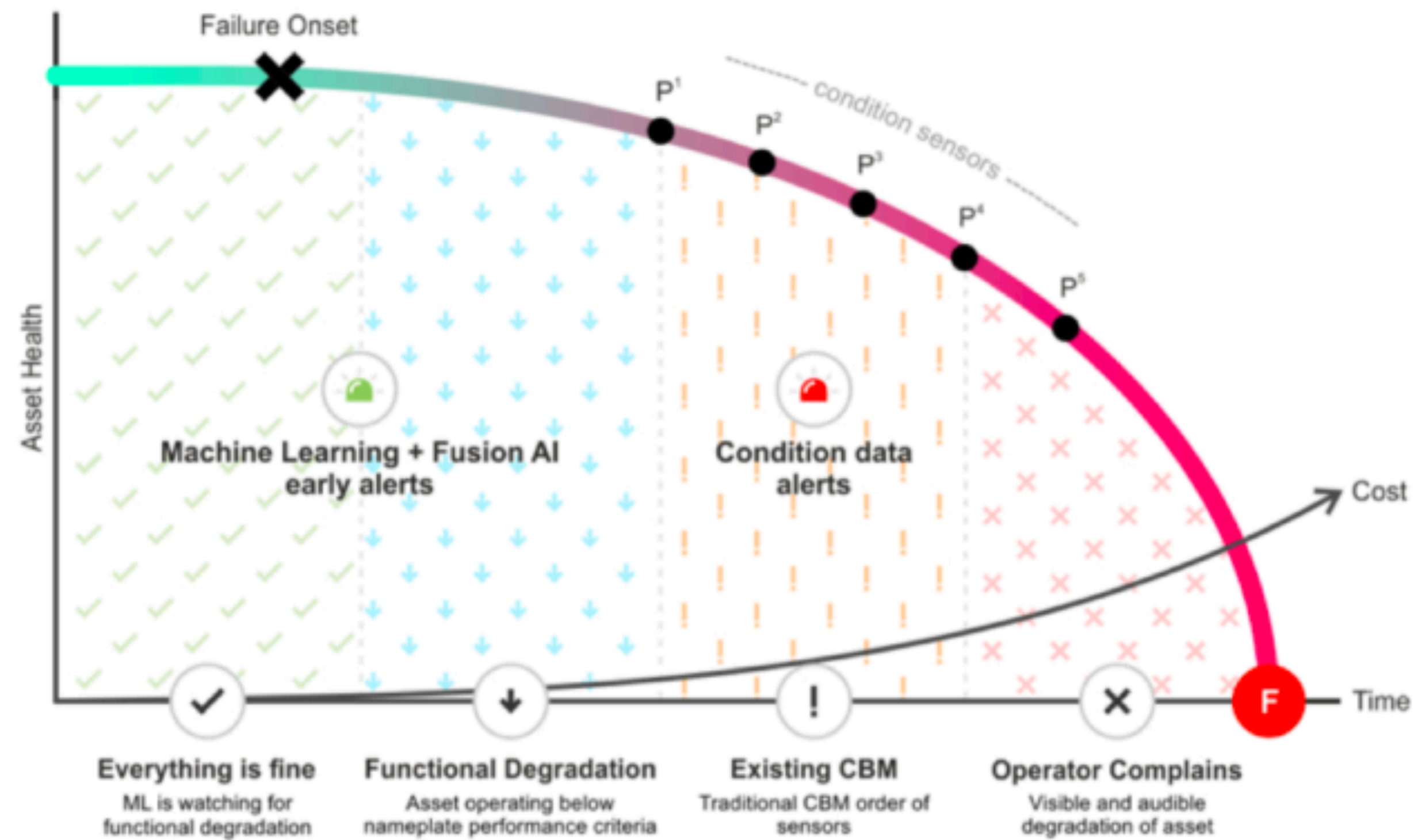
<https://www.gao.gov/products/gao-20-215sp>

Business of AI

Aggressive zero-data

P-F Curve 4.0

The role of Machine Learning



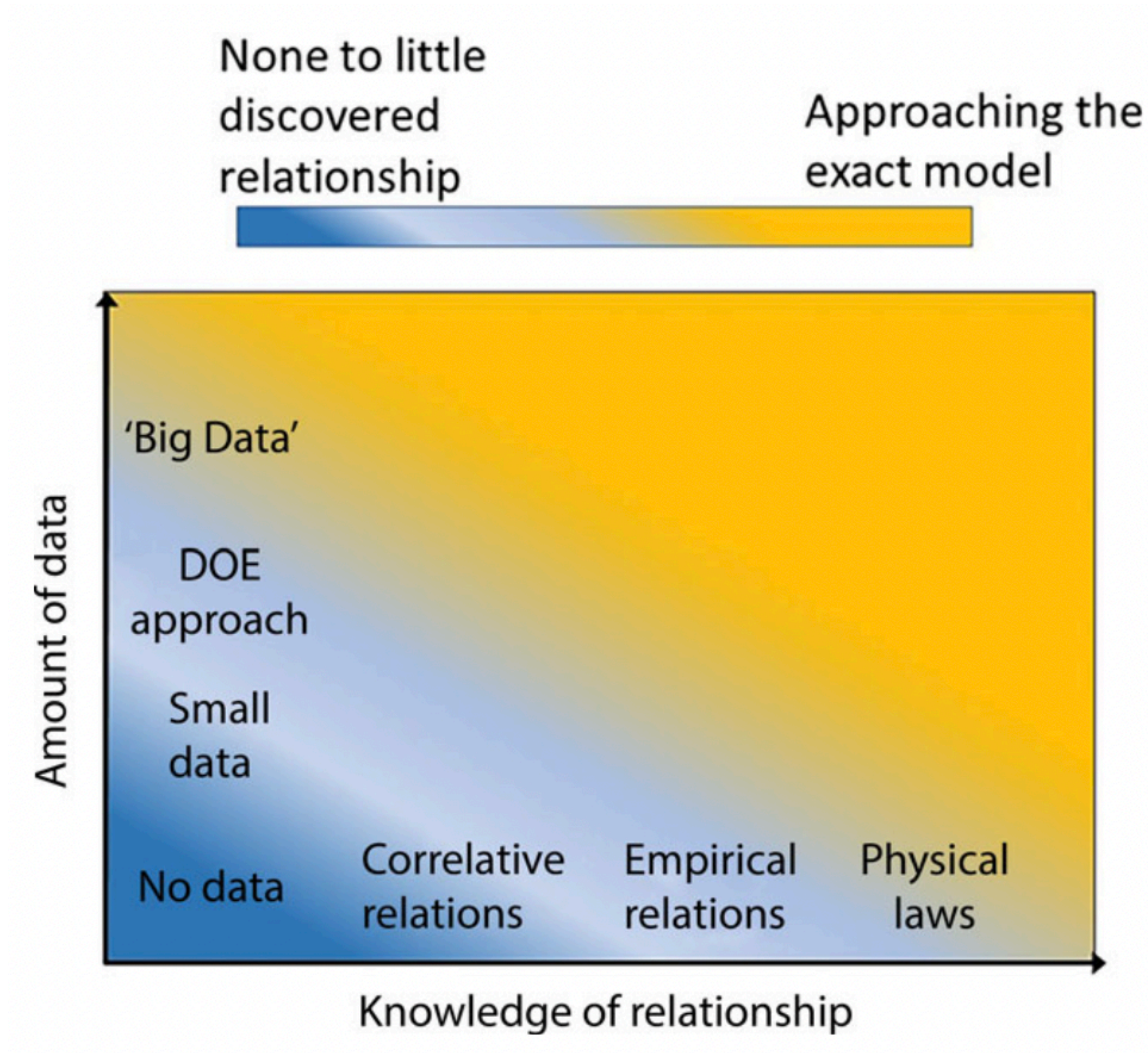
Business of AI

Aggressive zero-data

Category	Conservative big data	Aggressive zero-data
The “world”	Bits and bytes	Atoms and molecules
The process	Routine, often intuitive tasks (System 1)	Creative and scientific tasks (System 2)
Capitalization / moats	Massive data	Scientific insights
Unit economics	Linear / logarithmic improvements	Exponential improvements

Technology of AI

Aggressive zero-data



Embedding domain knowledge for machine learning of complex material systems


Technology of AI


Aggressive zero-data


Category	Conservative big data	Aggressive zero-data
Datasets size	Millions, billions samples	Hundreds, thousands samples
Targets and metrics	Single optimization criteria	Multi-criteria optimization
Mathematical problem	Pattern recognition	Extremum identification
Algorithms	Popular, commoditized, embedded in products (AWS, DataRobot, etc)	Physics-aware, extremely customized

People of AI

Aggressive zero-data




**Citrine Informatics**

**Stanford Graduate School of Business**

Greg Mulholland (He/Him) · 2nd


World Economic Forum Technology Pioneer, CEO, Board Member, and expert in digital transformation in Materials and Manufacturing





University of Cambridge
MPhil, Physics (Materials Science)
2007 – 2008
Activities and Societies: Trinity Hall, Trinity Hall Boat Club


- Thesis: Large-Scale Metal Oxide Nanowire Arrays for Hybrid Solar Cell Applications
- Research-based Master of Philosophy
- Supervised in Device Materials Group by Dr. Lukas Schmidt-Mende and Dr. Judith Driscoll
- Created anodized aluminum oxide templates for nanowire growth
- Fabricating Tin Dioxide and Titanium Dioxide nanorods using electrodeposition into template
- Nanostructures for use in hybrid organico-inorganic solar cell
- Attended courses in TEM, SEM, XRD, EDX, and electrochemical methods
- Regularly perform X-Ray Diffraction, Raman Spectroscopy, and Scanning Electron Microscopy for characterization of thin films and nanostructures
- Finalist, Gates–Cambridge Scholarship




INSILICO MEDICINE ACHIEVED PRECLINICAL CANDIDATE FOR NOVEL TARGET WITH NOVEL MOLECULE FOR A MAJOR DISEASE

AI-generated Novel molecule for the AI-discovered Novel biological target reaches preclinical candidate status in Idiopathic Pulmonary Fibrosis (IPF)

IN RECORD TIME, ON RECORD BUDGET WITH MINIMAL HUMAN INVOLVEMENT


**Insilico Medicine**

**Queen's University**

Alex (亞歷克斯 扎沃洛科夫) Zhavoronkov · 2nd

Founder and CEO at Insilico Medicine
Hong Kong SAR · [Contact info](#)

500+ connections





Lomonosov Moscow State University (MSU)
PhD, Physics

- Interaction of ions and chiral compounds in biological systems.
- Racemization of amino acids in proteins in the course of aging.



The Johns Hopkins University
Master of Science, Biotechnology

- Baltimore, Maryland, USA

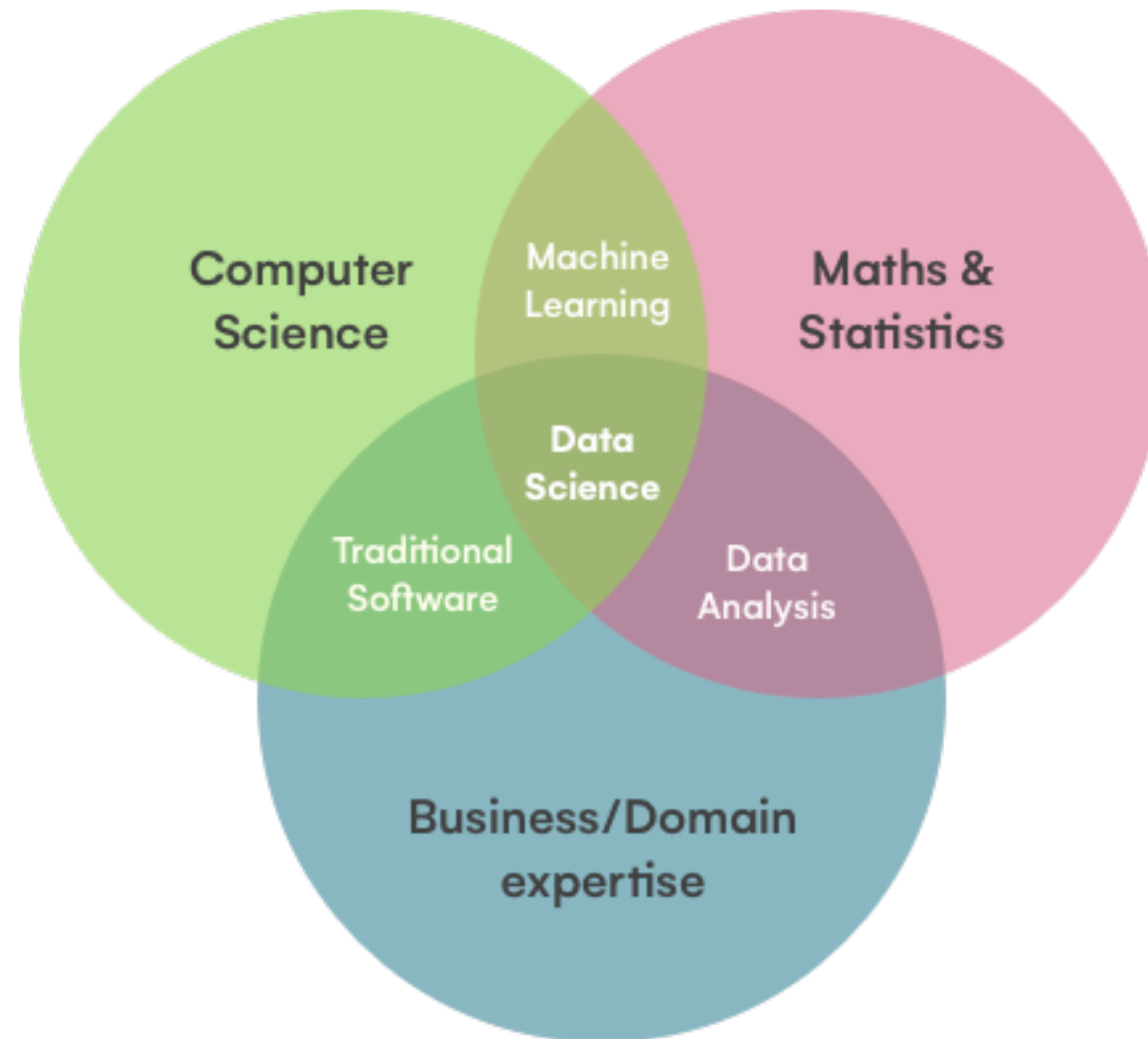
People of AI

Aggressive zero-data

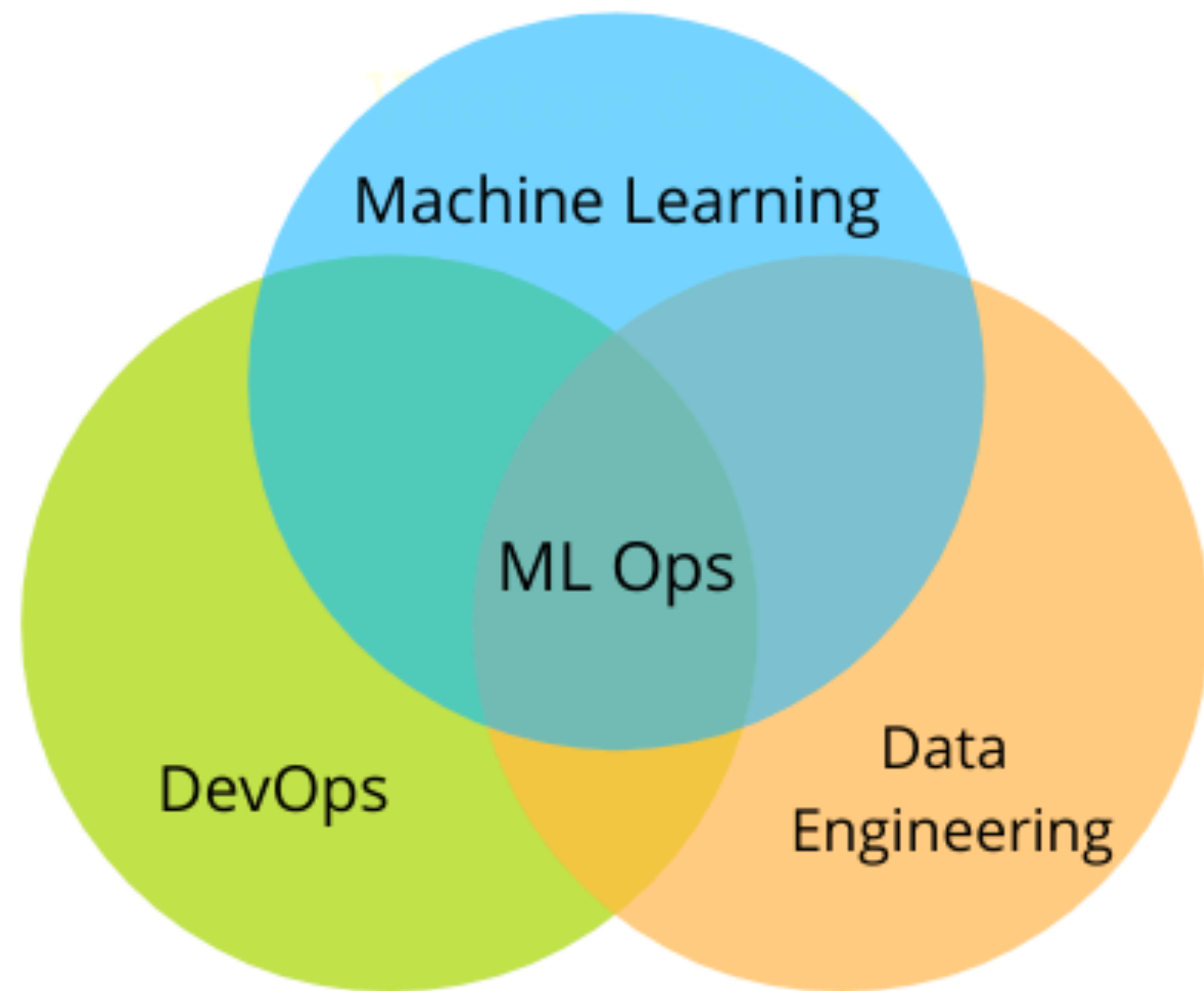
Category	Conservative big data	Aggressive zero-data
Stakeholders	Business, marketing	Subject matter experts, often with scientific background
Developers	Engineers-integrators	
Customers	“Regular users”	

Why anti-data-science?

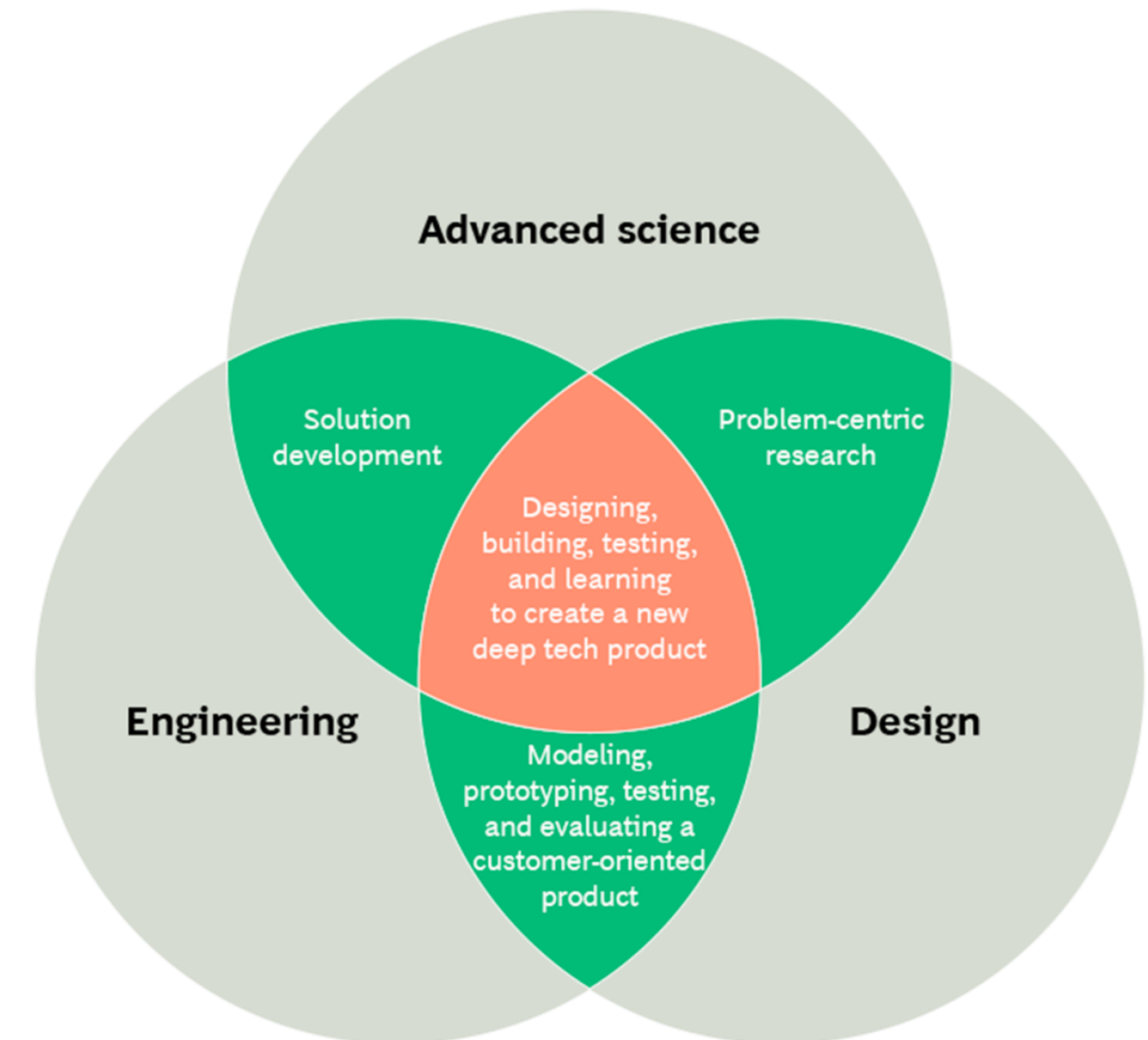
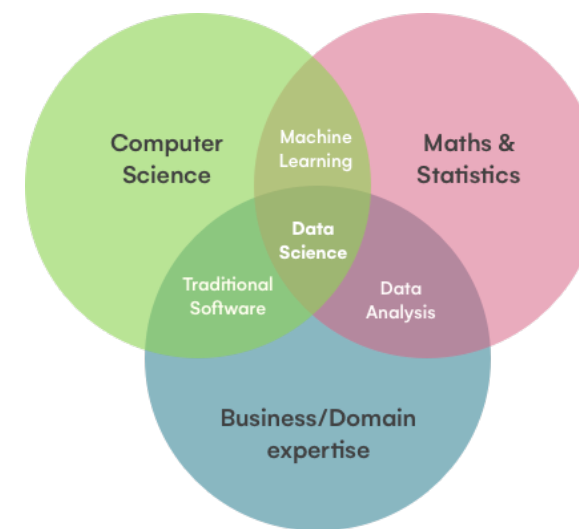
What is data science?



What is data science?



<https://neptune.ai/blog/end-to-end-mlops-platforms>

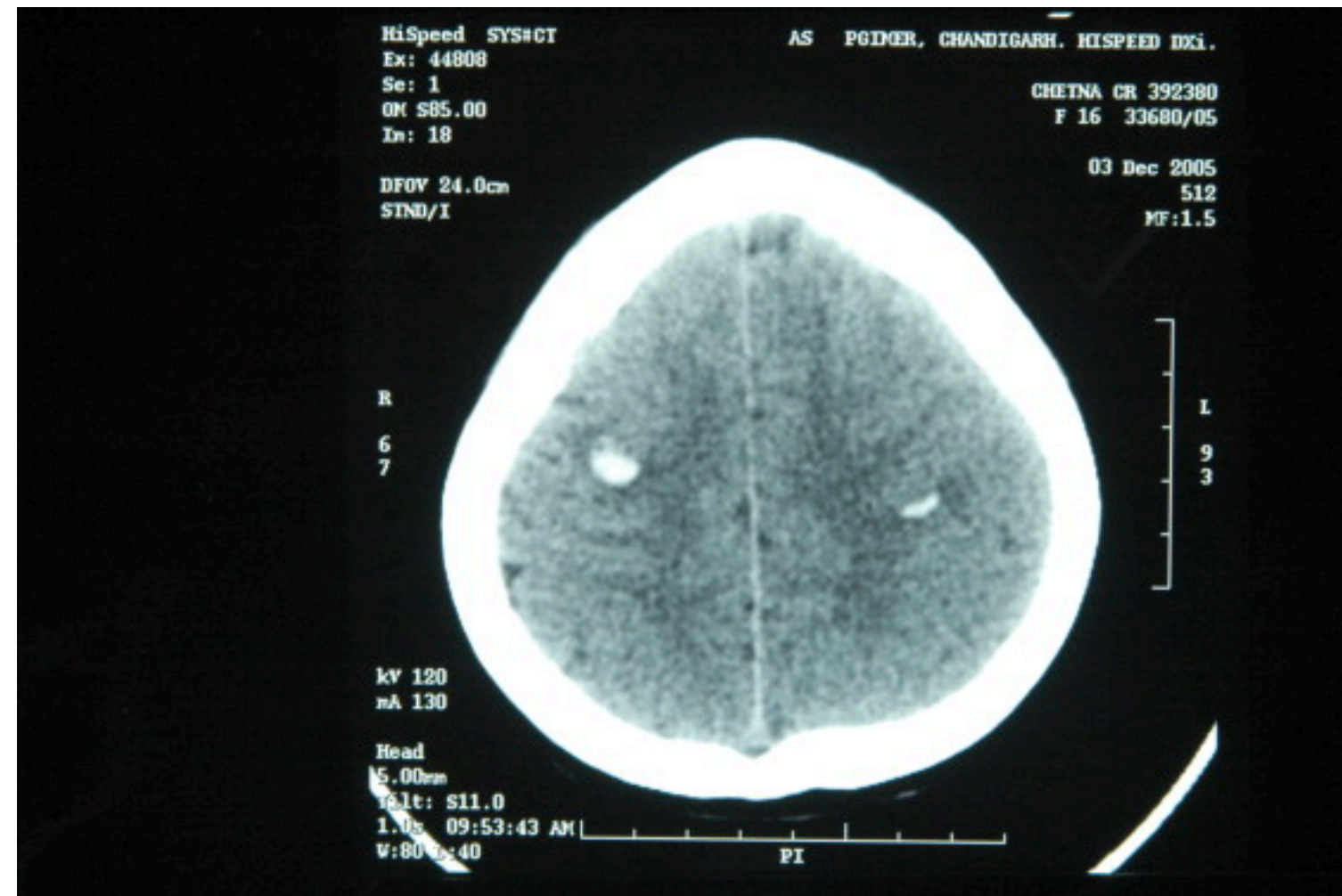


<https://www.bcg.com/publications/2021/deep-tech-innovation>

Anti-data-science

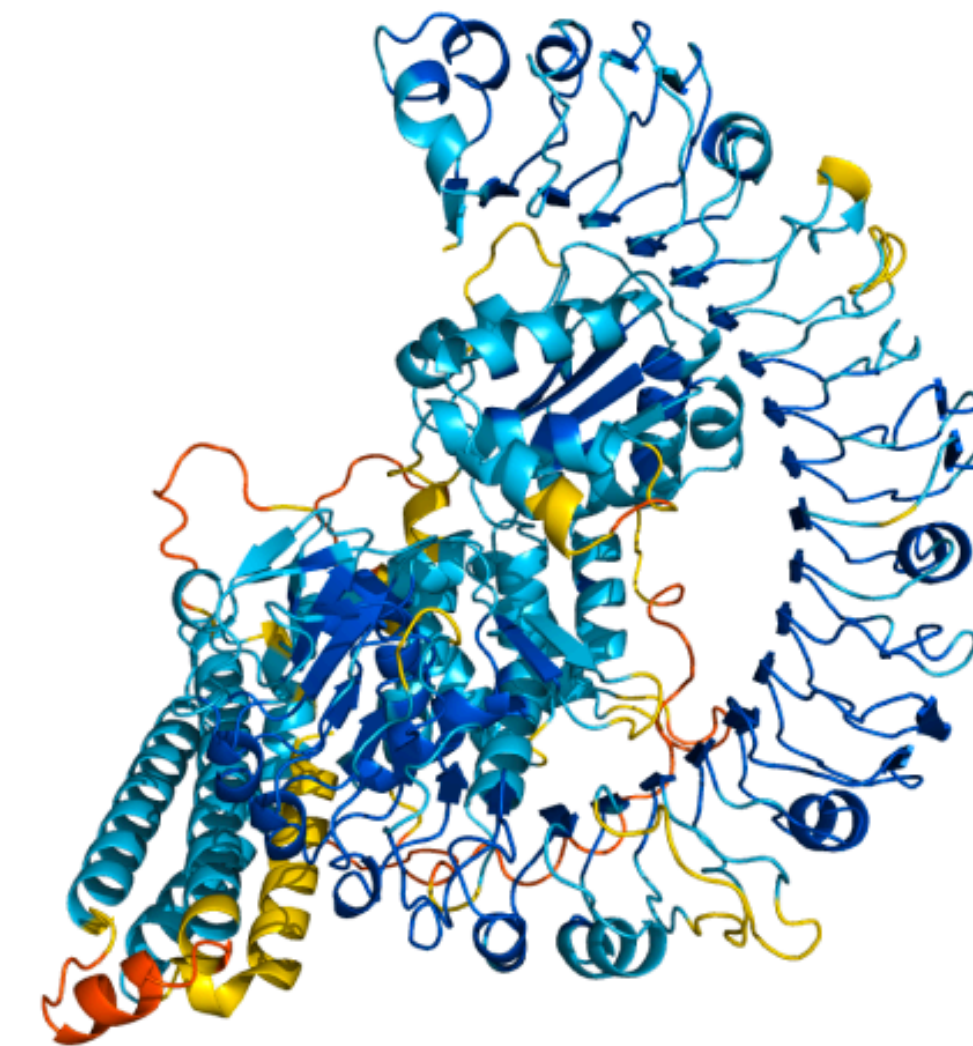
Applying “big data” tricks in biological world

🚫 Oncology, IBM Watson



- Trying to solve “routine” doctor tasks
- 53K samples, regular vision algorithms, focus on accuracy
- Organized by business people, done by engineers, doctors don’t use it

✅ AlphaFold, DeepMind



- Hard scientific task, novel finding
- 350K samples, satisfaction of chemical properties
- Done by scientists, with scientists, for scientists

Anti-data-science

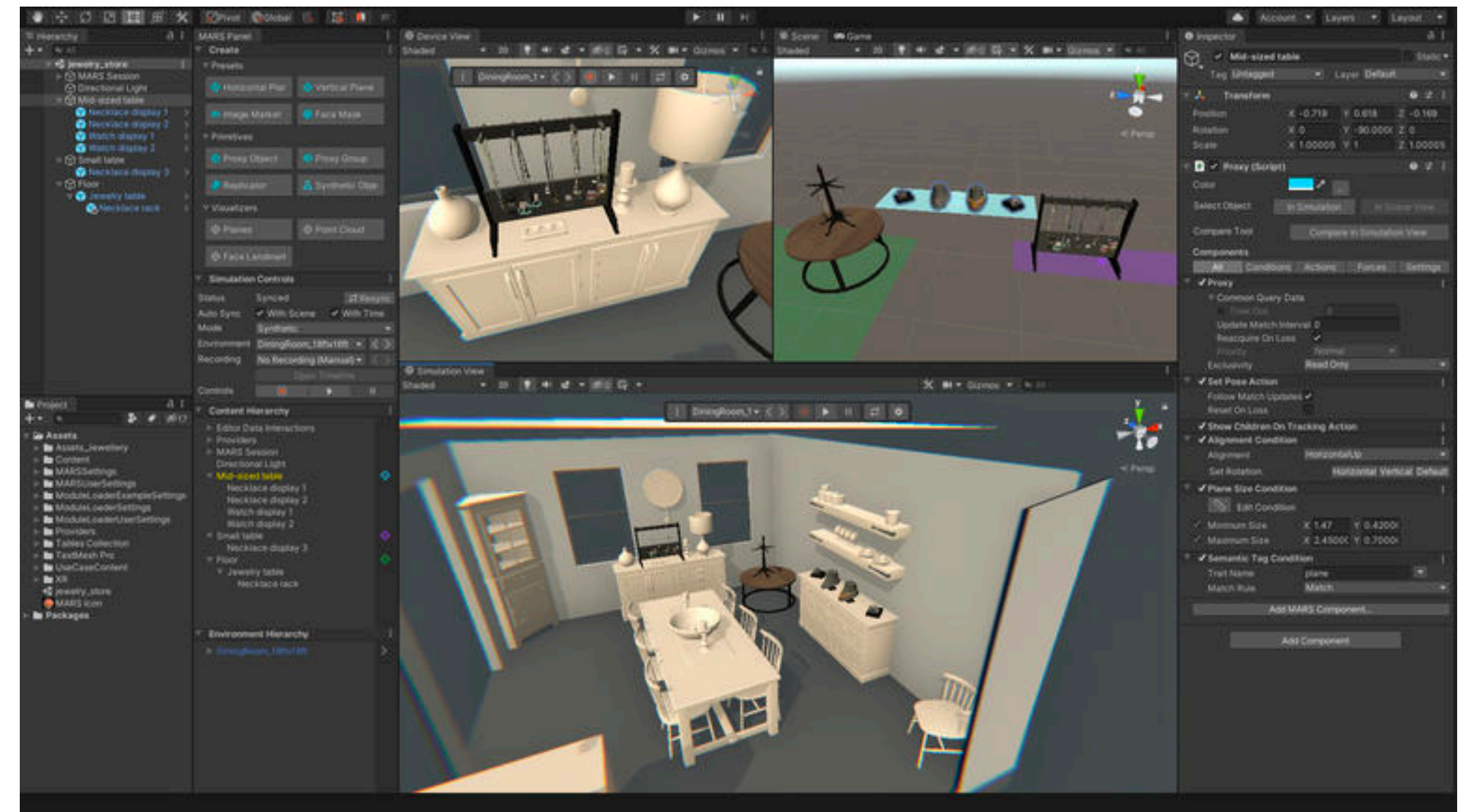
Applying “zero data” tricks in digital world

🚫 GANs for data augmentation



- Naive idea to use same old data to get “new data”
- Generating millions images with zero statistical value
- Organized by people that “just need more data”

✅ Unity, 3D modeling



- Requires expert knowledge in graphics and 3D
- You can generated limited high-quality samples
- Organized by domain experts with specific reqs



- Millions of people doing **digital routine tasks**
- Business driven, **linear** unit economics
- **Data monopoly** as a competitive edge
- Algorithms already **commoditized**

- Niche, scientific, often **physical challenges**
- Research driven, **exponential** unit economics
- **Expertise monopoly** as a competitive edge
- Algorithms **tailored** to the problem

let us help you to get on the right barbell side

find me in LinkedIn or ping us at neurons-lab.com