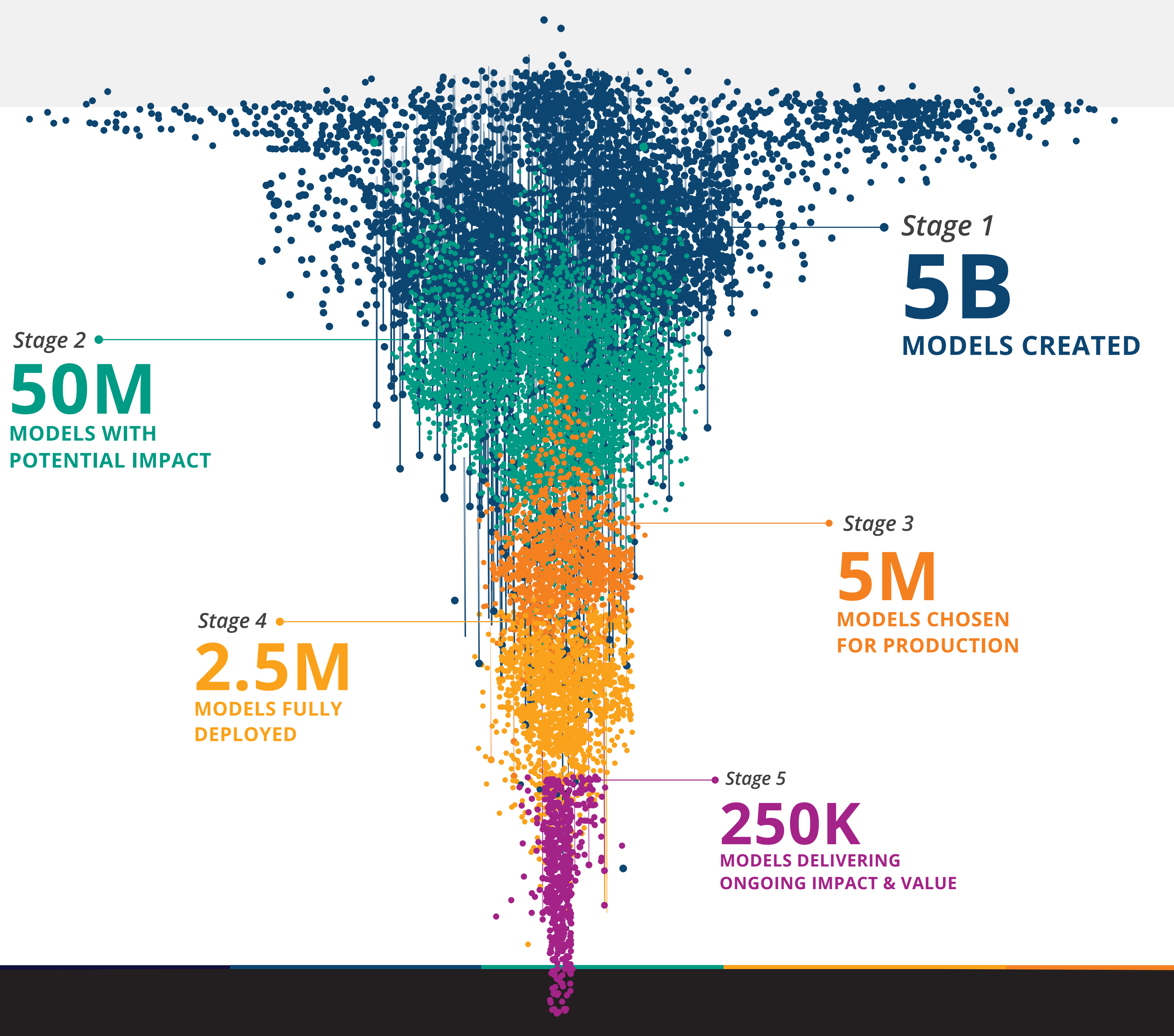
RAPIDMINER
An ALTAIR Company

The Model Impact Disaster

Less than 1% of models have their desired impact

Worldwide spending on AI is expected to grow by 19.6% in 2022, according to IDC. However, while spending is increasing, many organizations are still struggling to operationalize their machine learning models and maximize the investments they're making in data science.

RapidMiner's goal is to stop the disaster from spiraling out of control so that more models can make it into deployment and actually have an impact. How? By understanding the root causes of the disaster so we can develop concrete ways to alleviate them.



The 5 Stages of the Epidemic

5,000,000,000

Stage 1

CREATING MODELS

This is the starting point for every project, it requires you to prototype many models to find the right fit for your use case. Usually hundreds of models are created and evaluated during this stage per project.

$= 10^6 \times 5 \times 1000$

1M data scientists × use cases per project per year × model candidates per use case

50,000,000

Stage 2

FINDING MODELS WITH IMPACT

This is where you must evaluate, validate and test the hundreds of models to find the one that can actually deliver enhanced decision-making to drive cost reduction, revenue gain, or risk reduction.

$= \frac{5 \times 10^9}{1000}$

model candidates / model candidates per use case

5,000,000

Stage 3

CHOOSING MODELS FOR PRODUCTION

This is where the potential impact of the model is shared and discussed. As buy-in is garnered across stakeholders, resources are committed to deploy a model into production and timelines are put in place. Will this be an interactive dashboard? A weekly report? A part of a website? Estimates state that only 1 out of every 10 data science projects are chosen for production.

$= 50 \times 10^6 \times 10\%$

Models with impact × Models chosen for production

2,500,000

Stage 4

DEPLOYING MODELS

This is where the rubber hits the road and IT, DevOps and domain experts work to get the model fully operational within a business workflow. This often requires a diverse set of experience, including coding, computing resources management, and a deep understanding of the business workflow that's being enhanced. Estimates state that less than half of data science projects that are supposed to be in production end up fully deployed.

$= 5 \times 10^6 \times 50\%$

Models for production × Models fully deployed

250,000

Stage 5

DELIVERING ONGOING IMPACT AND VALUE

Once models are deployed the hard work is over, but the journey is not. It's still critical to monitor models for performance, business impact and potential drift issues. It may be necessary to change the actual model if alarming shifts occur. Continuous monitoring requires additional resources that many business don't account for, but is necessary to deliver ongoing impact.

$= 2.5 \times 10^6 \times 10\%$

Models in production × Models delivering ongoing value

Ultimately, less than 1% of models that can have an impact on a business are actually deployed & delivering value.

The Disaster is Making Headlines

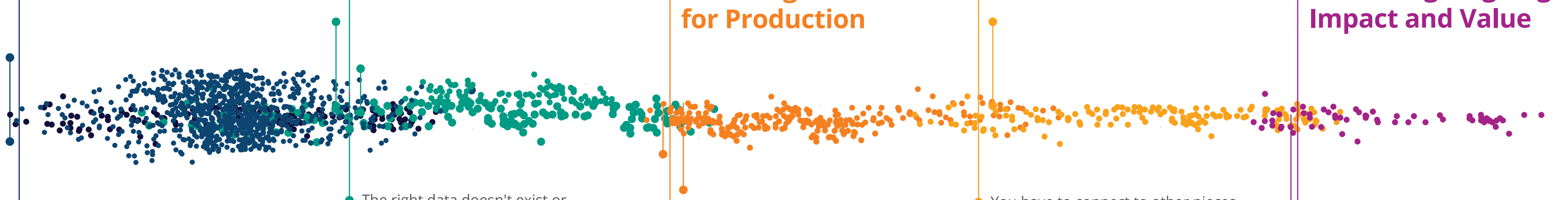
Extra, extra, read all about it! This troubling story is spreading like wildfire.

$85B *Expected worldwide AI spend this year*
- IDC

70% *of companies report minimal or no impact from AI*
- MIT

13% *of data science projects, or just one out of ten, actually make it to production*
– Venture Beat

Root Causes of the Disaster by Stage



Stage 1
Creating Models

- You must try 100s of models for every new project (No free lunch theorem)
- Auto ML reduces human effort, but still...
- ...this many models is hugely impacting energy expenditure
- If we use these resources we better make it count

Stage 2
Finding Models with Impact

- The right data doesn't exist or can't be accessed
- Rookie mistakes can be made by 'citizen data scientists'
- Models created without domain expertise may not make sense
- Data scientists can overinvest into solution paths and overlook simpler approaches (local optimum)
- Conformity bias – deep learning may not always be the best solution path

Stage 3
Choosing Models for Production

- Black box models
- Explaining models is hard and not always part of a data scientists core skillset
- May require separate analysis to show business impact, which is needed for buy-in
- Concerns over having the resources to deploy, monitor, and manage models
- Some decision-makers refuse to believe that models can trump intuition

Stage 4
Deploying Models

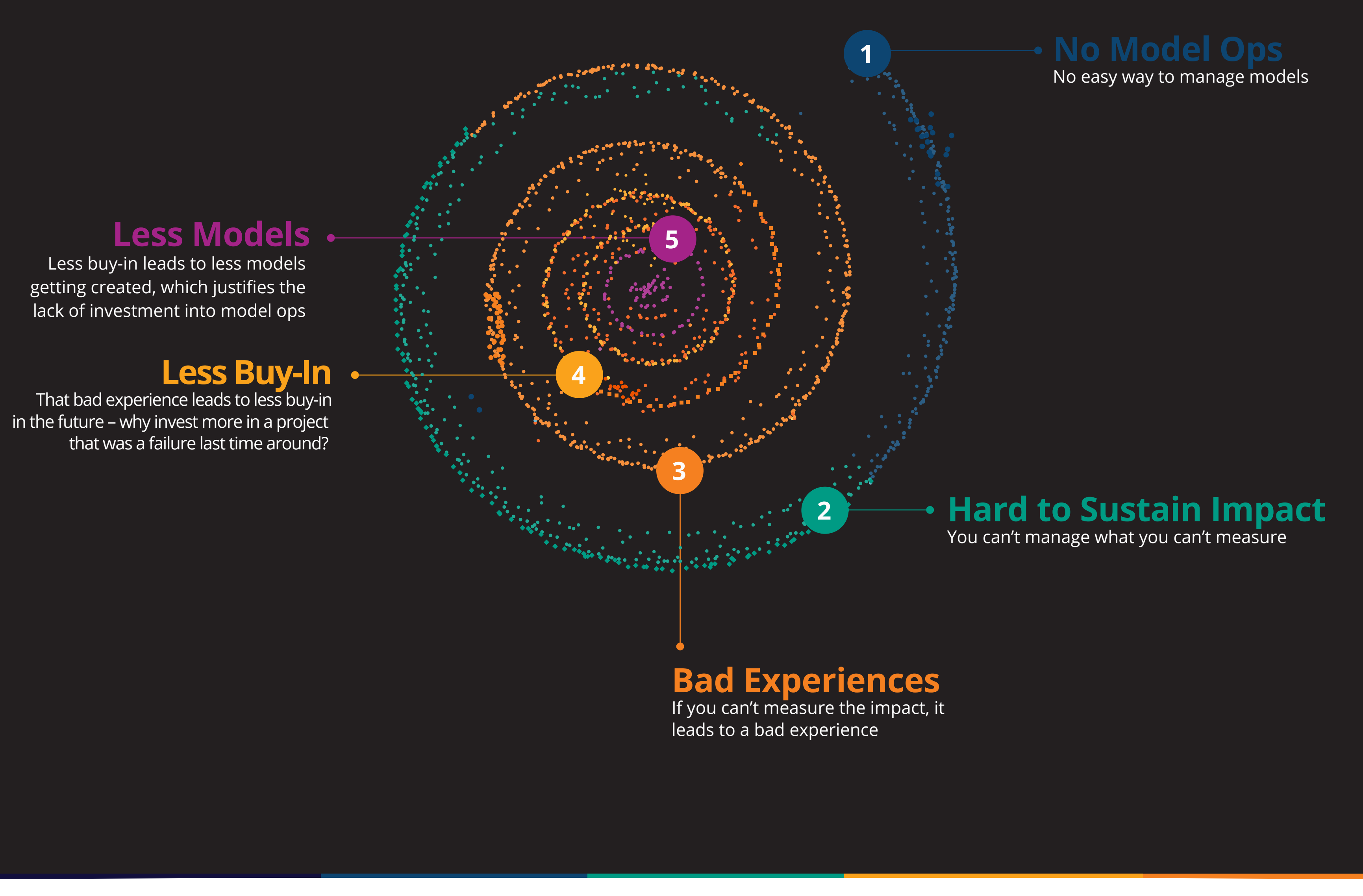
- You have to connect to other pieces of IT infrastructure, and you may not have permission
- You may need to rely on other people, and it may not be their priority
- Change requests mean you need to continue the perpetual convincing
- Creating proper solutions for model monitoring and management can be more work than building the model itself

Stage 5
Delivering Ongoing Impact and Value

- People don't understand the model or its predictions
- Need to do more convincing and explaining
- No ongoing monitoring of drift and bias
- Wrong validation in earlier stages may lead to less or even negative business impact
- Problems go undetected due to lack of model operations and management

The Spiral of Disillusionment

Some of the most widespread and avoidable issues that prevent good AI models from delivering the desired results occur between stages 4 & 5. This is because most organizations don't have sufficient model operations resources in place and it leads to bad experiences. We call this the 'Spiral of Disillusionment.'



1 No Model Ops
No easy way to manage models

2 Hard to Sustain Impact
You can't manage what you can't measure

3 Bad Experiences
If you can't measure the impact, it leads to a bad experience

4 Less Buy-In
That bad experience leads to less buy-in in the future – why invest more in a project that was a failure last time around?

5 Less Models
Less buy-in leads to less models getting created, which justifies the lack of investment into model ops

How RapidMiner Protects You from Disaster

Model Ops

Operationalize models faster and ensure long-term value creation without heavily relying on data scientists, developers, or your IT team.

- Reduce skill requirements to deploy worthy models
- Reduce risk by easily monitoring model performance and automatically guarding against drift & bias
- Deliver insights using custom dashboards, AI apps, or your preferred BI tool

Value-Sensitive Scoring

Assess and share the financial impact of your models, no matter the use case.

- Go beyond traditional accuracy stats to show your models' cumulative gains and financial impact
- Analyze scoring times to pick models that are fast enough for your desired use case
- Compare distribution differences between predictions and actual values

AI App Building

Put powerful insights into the hands of decision-makers in an engaging and intuitive way.

- Connect and share model results in consumable AI apps without writing a line of code
- Expose simulations, evaluate results against expectations, and run "what-if" scenarios to better understand predictions
- Run apps on any device including tablets and mobile phones to support your flexible workforce

Built-In Trust & Transparency

Understand how models make predictions and show your work at every project step to build confidence in promising AI solutions.

- Master the entire data pipeline with automatic process documentation in visual workflows
- Calculate global and local feature weights to understand the driving factors for models' predictions and discern each prediction's most influential features

Fits with Existing Enterprise Systems & Architecture

Simplify the process of integrating disparate data sources and delivering insights where they're easiest to consume.

- Connect to today's most popular enterprise applications and ingest relevant data no matter its size, format, or velocity
- Use RapidMiner's containerized architecture to avoid deployment friction and an overreliance on developer & IT resources

RAPIDMINER
An ALTAIR Company

For those driven to accelerate the pace of transformation, RapidMiner is the enterprise-ready data science platform that amplifies the collective impact of your people, expertise, and data for break-through competitive advantage. RapidMiner's data science platform supports all analytics users across the full AI lifecycle. The RapidMiner Academy and Center of Excellence methodology ensure customers are successful, no matter their experience or resource levels. Since 2007, more than 1 million professionals and 40,000 organizations in over 150 countries have relied on RapidMiner to bring data science closer to their business.

rapidminer.com

IDC, IDC Forecasts Companies to Increase Spend on AI Solutions by 19.6% in 2022, 15 Feb 2022.
Venture Beat, Why do 87% of data science projects never make it into production? 19 July 2019.
MIT, Winning with AI, Sam Ransbotham, Shervin Khodabandeh, Ronny Fehling, Burt LaFountain, and David Kiron, 15 Oct 2019.