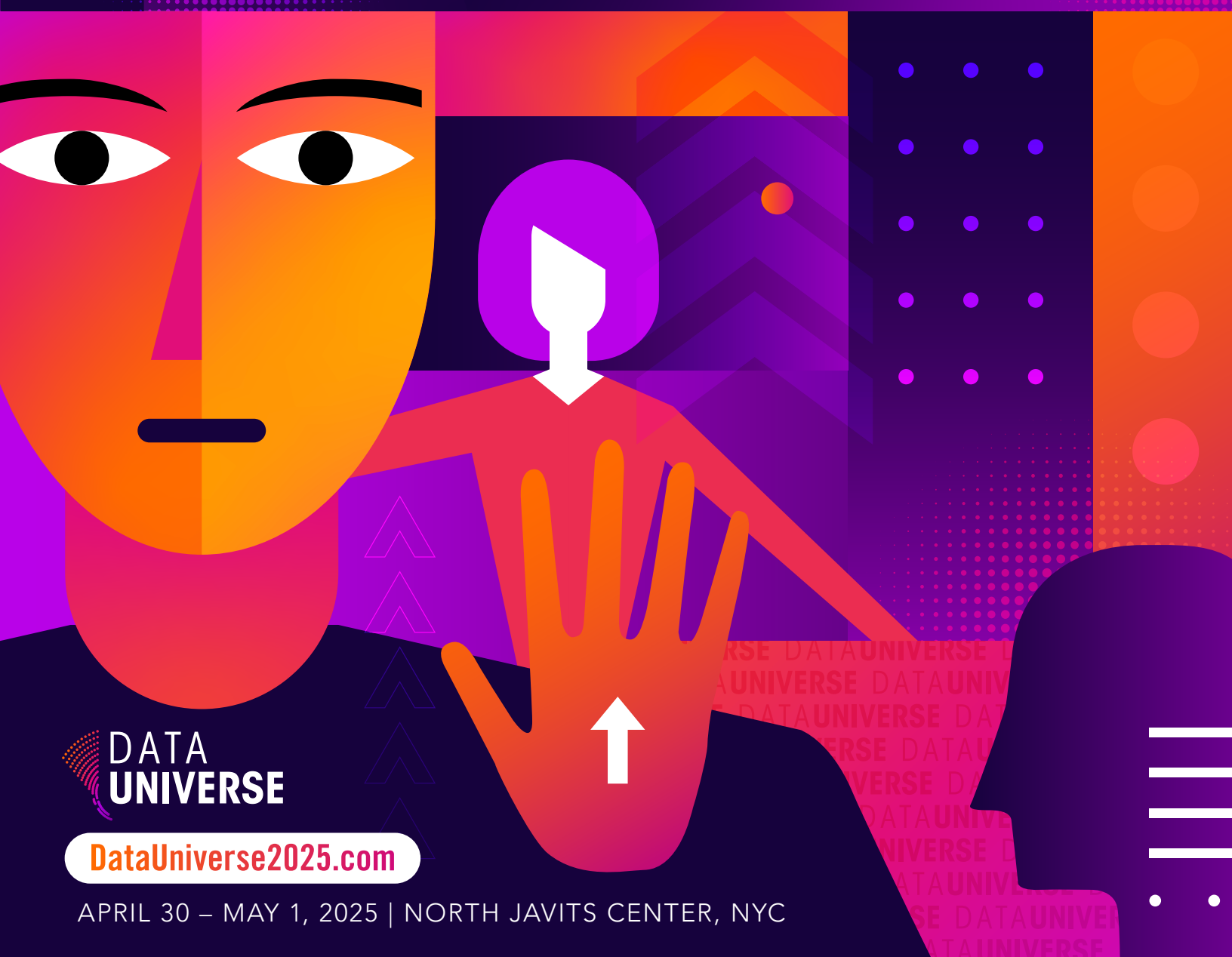Data Universe 2025 Manifesto

Decide Better.

BY ALISTAIR CROLL

DATA UNIVERSE

DataUniverse2025.com

APRIL 30 – MAY 1, 2025 | NORTH JAVITS CENTER, NYC

Data Universe 2025 Manifesto

Decide Better.

BY ALISTAIR CROLL

All of data science can be summarized in two words: *Decide better.* But decision without action is simply dreaming. So what compels us to act?

At an individual level, we have some ideas. In the late 1800s, mass production and industrialization were creating a new wealthy class. While poverty and racism were still widespread, more workers had discretionary income. In much of Western society, our fundamental needs were met, with enough left over for *desires.*

Ad-man Elias St. Elmo Lewis described how our decisions take us from attention to action in the Attention, Information, Desire, Action (AIDA) model which tries to explain how a consumer goes from initial awareness to a change in behavior:

1. Something captures our **attention;**

2. We collect **information** about it;

3. We discover a gap between our current state and an imagined, better stage, triggering **desire**;

4. We **act**.

But if deciding to act is so straightforward, why are we so often paralyzed?

Why we can't act

A century after Lewis first proposed his model, we now know that turning decisions into actions is far more complicated than he thought.

At an individual level, perfect decisions take a lot of time and effort. Instead, we settle for "good enough." There's even a word for this: *Satisficing*—a portmanteau of *satisfy* and *suffice*—was coined by Nobel-winning economist Herbert Simon. And that's when we're individuals trying to act rationally.

Compelling groups to act is far more complicated, because we face:

- **Hidden motivations**
  Everyone has their own agenda, which they may not share.

- **Groupthink**
  We can't read others' minds, so we act according to what we think others think.

- **A desire to fit in**
  We're social, irrational creatures who value belonging more than we do being correct.

- **Coordination problems**
  When we compete, a lack of trust leads to suboptimal outcomes.

Let's look at these four obstacles in more detail.

Hidden Motivations

If you ask several people to think of a car, it's unlikely they'll all imagine the same make, model, and color. Our brains fill in the missing details, and they're different. When working with data, the same information leads people to different conclusions.

First of all, we gather different information based on our individual senses, our own biases, and the past experience we bring to a decision. Our desires include not just those created by that information, but also our experience, context, and expertise. We decide what queries to run, how to sort tables, what to filter, and what to discuss with others.

The data we collect is also shaped by our own motivations. Someone who's afraid of losing their job wants safety and predictability, while someone who's exhausted just wants the meeting to end, which affects how they interpret information. We're often in unspoken, even unconscious, competition with our colleagues, each of whom have their own goals and needs.

As a result, even a team that starts with the same data can reach different conclusions, and act in very different ways.

Groupthink

A 2022 study showed that 73% of Americans—a sizable majority—support meaningful action on climate change. But when researchers asked those same people what percent of their fellow Americans agreed with them, the number was a paltry 40%.

How can we be so wrong?

Part of the answer lies in the disproportionate attention we pay to vocal minorities. When millions of people see a film, they seldom form picket lines to let the world know. Yet a small group of objectors can easily make the evening news. Media and social platforms amplify outrage, pushing their messages into our feeds until everyone believes that a tiny, loud group represents the majority.

Across topics as disparate as flat-earther conspiracies, science-denying wellness experts, gun rights absolutists, and startup Founder Mode hustle culture, we're terrible at understanding what others are really thinking, even when we have hard data to guide us.

Social Death

Humans are social creatures. Evolutionarily, there's safety in numbers, and working together lets us accomplish amazing things.

To ensure that we contribute and don't take more than our fair share, we punish those who don't pitch in with ostracism. As 2024 Data Universe keynote speaker David McRaney reminded us, "social death is greater than physical death." We're terrified of speaking out and being shamed. We often know a decision is wrong, but say nothing. It's easier to remain silent—and not be implicated in a potential failure—than to speak up and suffer the political consequences.

Even when we know the data is wrong or the action is questionable, we often keep it to ourselves.

Coordination Problems

Game theory is the study of how competing decision-makers interact. The simplest competitions are zero-sum games with a clear winner and loser. The players are opponents, and their competing interests are obvious. Such systems settle down fairly quickly: If most people decide to drive on the same side of the road, the risks of doing otherwise become significant.

Most competitive interactions are far more complex, and may even benefit everyone involved. But this requires trust. Consider a group of cavemen deciding what to hunt. If they cooperate, they can take down a mammoth and divide the spoils. Acting alone, they'll only hunt smaller creatures. Each hunter could cheat, avoiding real risk, and still benefit—but if every hunter cheats, they won't catch their prey. In these more complex games, every participant's actions impact the outcomes of others, so trust becomes important.

Even when participants trust and communicate, unexpected side effects of their decisions can become clear only long after they've acted:

- Heavy trucks damage roads more than light cars do, but both benefit from the road.
- Single-use plastics were convenient, but have littered the oceans and filled our bodies with dangerous microplastics.
- A frequent and an occasional driver pay the same monthly insurance, but the occasional driver unfairly subsidizes the frequent driver's insurance.
- Burning fossil fuels more quickly than their carbon can be sequestered brought us the advantages of the industrial age, but has forever changed the planet's climate.

These externalities, not apparent to participants when decisions are made, often result in regulation and complexity later on. Stripe founder Patrick Collison keeps a list of things that were accomplished quickly in the past, contrasting them with beleaguered projects that spend much of their time navigating regulatory environments.

Game theory, imperfect information, and unanticipated outcomes are symptoms of coordination problems. These occur when the perverse incentives of a system make everyone act against their own best interests. For example, no country wants to spend money on a standing army—but everyone has an army because everyone else has an army. Science explainer Liv Boeree points to the use of filters on social platforms as a coordination trap: Everyone uses them, changing everyone's idea of beauty, which prompts everyone to use them.

Taken together, our hidden agendas, misunderstanding of what others are thinking, unwillingness to speak out, and the perverse incentives of the systems we're in mean that even the best data often leads to bad outcomes.

So how do we fix it?

What Everyone Knows Everyone Knows

In the story of the Emperor's New Clothes, two con artists scam a vain emperor. They tell him they can craft a fabric so fine, it's almost unnoticeable; and that only the wisest of people can see the clothes they make. They also set up a trap: To question the clothes is to admit stupidity, and nobody wants to look stupid. In some versions of the tale, the emperor threatens to behead anyone who questions his marvelous vestments.

The courtiers play along. They ignore what their senses tell them, and compliment the king on his fine robes. It's only when a young child blurts out that the emperor is naked that everyone else finally acknowledges the lie.

The tale is far more than a children's story. It's an illustration of how groups of people—from businesses to political parties to societies as a whole—make decisions. Because the antidote to coordination problems is common knowledge.

When everyone has access to a piece of information, but sharing that information comes at a high cost, we must turn it into common knowledge if we want everyone to act on it. Once *I know that you know that I know*, our social muscles propel us forward. In fact, it doesn't even have to be everyone—studies show that as few as 25% of a group can put forward a change around which the rest of the group quickly coalesces.

From Data to Outcome

While the role of data is to help us decide better, the *goal* of data is to produce better outcomes. It's not enough to collect, process, and share data well. We must create the conditions by which that data leads to actions. Every data-ready organization needs to align its teams with clear messages, calling out hidden motivations. It must better understand what is opposing the change that the data suggests, and the popularity of the ideas against which it competes. It must encourage plainspoken criticism, removing the thread of ostracism. And it must turn data into common knowledge, so that everyone knows everyone knows the facts.

Data needs to not only decide better, but make us act better. It must be the child at the Emperor's parade, willing to speak truths so self-evident in hindsight that suddenly, *we know everyone knows*. This is a challenge not only of technology, but of organizational culture, governance, provenance, and leadership.

The Data Universe Imperative

Data Universe 2025 casts a wide net on the topic of data, asking not only where the rapid pace of innovation is leading us, but also how to harness that change in service of the only thing that really matters: Better outcomes.

By bringing together experts from technology, business, and academia, Data Universe is a forum where we can collectively navigate the complexities of our data-driven world. We invite you to join us in this crucial conversation, as we work together to not just decide better, but to act smarter in the face of unprecedented challenges and opportunities.

SOME HOT TOPICS WE'LL BE FOCUSING ON THIS YEAR

The Second Order Effects of Generative AI

Even if AI advances were to stop today, it would take many years for enterprises to deploy them. Many of the processes on which we rely assume that certain activities are easy, while others are hard.

- **WHEN HARD GETS EASY:** Writing an application for a grant, for example, required considerable time and effort—so grant applications were carefully read by experts. Now that an LLM can generate a proposal in minutes, how should the granting process change?

- **WHEN EASY GETS HARD:** Walk into a bank circa 1970, and the manager recognizes you by name.

Today, you're merely an account number. Scammers, armed with all of their victim's personal information, can often seem more credible than those they defraud. And generative AI can now mimic voice, image, and other biometric details. Now that AI undermines trust, how do we decide what is true?

Across every facet of a business, operators need to reconsider what is now easy and what is now hard, and adjust their processes accordingly to mitigate the second-order consequences of AI.

The Rise of Shadow IT

At the dawn of the web, Software-as-a-Service startups infiltrated traditional IT departments by making it trivially easy for anyone with a web browser to use their product. Before that time, software came from a central authority, using client-server architectures, often installed from floppy disks or CD-ROM.

It wasn't long until CFOs were analyzing their companies' credit card statements, discovering that they were unwittingly paying companies like Salesforce and Amazon Web Services for a second, parallel, IT department run by rogue employees. This was known as shadow IT.

Statista estimates that there are 28.7M software developers in the world today. There are vastly more "knowledge workers" who have access to a computer and work with information. Recent advances in AI promise to lower the barrier to software development, allowing many of these people to write personal, transient applications. Consider the teacher building a testing tool, or the doctor writing a simple app to help with diagnosis, or the lawyer coding a billing system.

Now consider the legal and ethical implications of that software. It has no lifecycle management, no testing program, no version control. It may be unsupportable. A second wave of Shadow IT is about to break, cresting not because of the easy deployment cloud computing delivered, but because of the easy development AI—supported by turnkey compute platforms—promises.

New Kinds of Data Storage

There are many kinds of database, each optimized for a specific task. Relational databases were great for linking many tables of information; time-series databases worked well with sequential information. We developed special databases for social graphs, genomics, and more. We poured disparate data into data lakes and warehouses in the hope of one day extracting competitive advantage.

Despite this, the bulk of data work still involves cleaning, labeling, and structuring information.

Research in adjacent fields is bringing us new kinds of databases, from those used to store large language models, to social graphs, to implementations of search as a database. These more esoteric data systems are poised to move into the mainstream, reducing complexity and offering new ways to analyze information quickly and cost-effectively.

Can Governance and Performance Coexist?

Everything has pros and cons, which is why, it's said, that the opposite of progress is congress.

Careful stewards of information set up elaborate policies to prevent its mis-use, seeking consent and approval.

They implement encryption, privacy policies, and disclosure. Meanwhile, reckless competitors throw caution to the wind, running roughshod over data rights as they capture market share.

Regulation alone won't fix this problem, particularly when the services on which we rely cross international borders. So how can a company be a good actor and still win in a competitive marketplace? How do we create incentives that safeguard our most precious and personal information while punishing those who ignore the law?

Data Literacy and Democratization

Improving data literacy across all levels of an organization is crucial. However, this must be balanced with robust data governance to ensure data quality and trustworthiness. The tension between making data widely accessible and maintaining control and security is a key challenge.

Emerging Technologies and Data Ecosystems

Emerging technologies like edge computing, quantum computing, and blockchain are reshaping the data landscape. The growing importance of data ecosystems, marketplaces, licensing and partnerships is in sharp relief with Billions of dollars of lawsuits threatening generative AI. While our focus is on data and outcomes, we need to understand how other technologies and business models let us put data to work.

The Future of Work

The evolving data landscape is fundamentally changing job roles and required skills in organizations. We haven't even answered basic questions. Consider, for example, an employee who automates their work entirely and continues to collect a paycheck. Should they be rewarded for their ingenuity, and urged to continue automating other parts of the business? Or should they be fired for not doing their job?

Technology has replaced individual job functions for centuries. Nobody bemoans the loss of lamplighters that came with the electrification of cities. But agentic AI promises to replace not a single job, but an entire department. When algorithms prepare and send business proposals to clients, what happens to the proposal-writers? When diffusion tools supplant concept art, what happens to the designers? When a startup's technical co-founder is a mashup of an LLM, a development environment like Cursor, and a runtime platform like Replit, how does venture capital adapt?

We can't foresee the seismic changes that the very nature of work is about to undergo—but we can at least think about the important questions to ask.