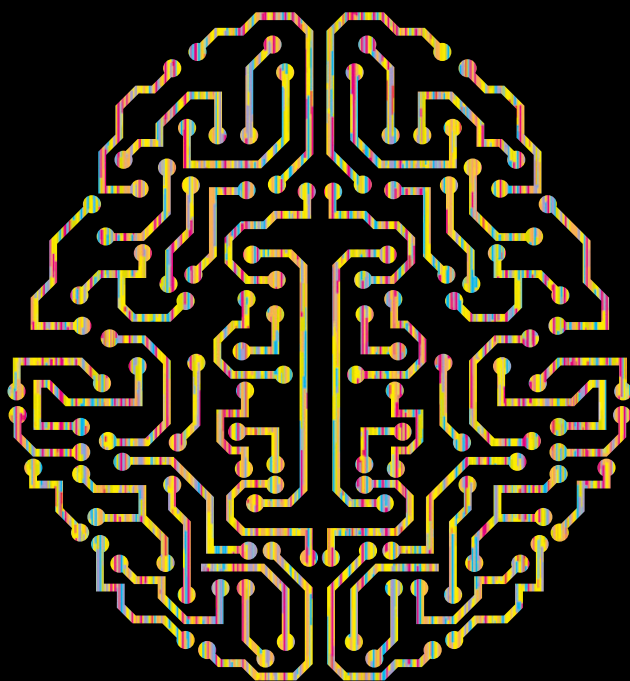# THE WORLD OF DATA

—

Between clarity and white noise

SCENARIOS FOR THE DATA SOCIETY

It has sparked a wave of euphoria in research and business. A huge advance has been set in motion: machines are starting to think for themselves. After decades of collecting and storing data on the world, we now finally have the means to connect and process all this information. Complex algorithms are bringing hitherto inert data to life and influencing the lives of everyone – from Google search results to the discovery of the Higgs particle or even the stock market crash. But anyone who understands data knows we cannot leave thinking to computers alone. The more intelligent the machines become, the more we have to use our own minds. For it will always be up to human beings to decide what makes sense.

I. MEASURING

"Datafication" is the buzzword experts use for the process of turning everything into data. The year 2011 saw more than 2 zettabytes of new data being generated worldwide, and studies estimate that a further 2.7 zettabytes of data were added in 2012. McKinsey & Co. forecast that this volume will increase by 40% annually until 2020. The amount of data being produced is growing faster than storage capacities.

WHAT WE THINK:
2.9 million emails are sent every second worldwide; every minute sees 660,000 new posts on Facebook.

WHAT WE FEEL:
Some 35,000 individual likes per minute on Facebook and countless emoticons in text messages reveal our current emotional states.

WHERE WE ARE:
GPS technology in mobile phones tracks our movements; there are more than 2,100 check-ins per minute on the location-based social networking website Foursquare alone.

! At the start of 2012, only 2% of all data was not stored digitally.

! 90% of all the data in the world was generated in the last two years

! By 2050, every single person in the world – with its estimated total population of nine billion – will have produced an average of 489 terabytes of personal data.

WHAT WE BUY:
Retailers and payment systems such as Paypal or Visa document every single transaction. Apple records 47,000 app downloads per minute.

WHAT WE WATCH:
Every minute we upload 48 hours of new video material to YouTube and post around 7,000 photos on Flickr and Instagram.

WHAT WE LOOK FOR:
No fewer than two million search queries per minute are processed by Google alone – not to mention all the searches on Baidu, Yahoo, Bing etc.

II. CONNECTING

It is connecting all the disparate information that is the Achilles' heel of the data world hype. Useful information is often acquired only when various data sets are put together – for instance when the combination of the locations and speeds of cars shows where there is a traffic jam. Companies such as Acxiom, Google or Facebook as well as intelligence services are in possession of particularly large volumes of data.

! Only 15% of all data is structured and can be used in sortable tables. The rest consists of data such as emails or eBay transactions.

! Most data sets cannot be combined with one another for legal reasons such as data protection or license restrictions.

! In order to make it possible the "Internet of things", in which machines exchange data automatically, almost 340 sextillion new web addresses are being created.

III. PREDICTING

Life in the data society is changing the way we see the world. Even the discovery of the Higgs particle was based on the analysis of huge amounts of data on physical measurements. However, in essence, the analysis of large volumes of data is a matter of prediction. These forecasts are not based on chains of cause and effect, as would be the case in human reasoning, but on algorithms, i.e. mathematical processes seeking patterns among the data masses. The end result is probabilities – for example what developments we are likely to see on the price front or whether that guy who seems to be in such a hurry is about to rob a bank.

! We all use the results of the processing of large data sets on a daily basis. They are behind spam filters, online dating services and the Autocorrect function in mobile phones.

! Algorithmic predictions are contingent on the continuing general availability of personal data.

! There are only around 400 known statistical regression models, i.e. processes with which to find systematic connections among data sets.

FLASH CRASH
Sudden drop in securities prices on the stock market on May 6, 2010, when a flawed algorithm wiped out 6% of the value of the America's S&P 500 index within minutes.

BURNING BRONX
An attempt in the 1970s to use mathematical formulae to find the ideal location for fire stations in New York led to frequent fires in the South Bronx which destroyed 40% of the buildings in the area.

EAVESDROPPING ON THE CHANCELLOR
Automated US surveillance programs tapped the phones of the heads of government in Germany and Brazil, triggering major diplomatic incidents.

MEDICAL DIAGNOSES
It wasn't doctors but a software program that discovered when processing large sets of data that babies with a regular heartbeat are more susceptible to illness than those with an irregular heartbeat.

CRIME FIGHTING
Reduction in crime due to data-based risk assessment aimed at helping police deploy resources to maximum effect in order to prevent crime.

PREDICTIONS
A world simulator being worked on at the ETH in Zurich – a sort of data-driven Internet on which anyone can use the search engine like Google to ask questions about the current state of the world and probable future developments – is intended one day to help predict economic crises and wars, and to prevent them.

ROBOT JOURNALISTS
Artificial intelligence specialists Narrative Science have developed algorithms that filter information from statistics or Twitter feeds as the basis of very simple reports about share prices or local football matches.

IMPROVED DISASTER RELIEF
The Fraunhofer Institute is currently developing an instrument with which to optimise disaster relief deliveries – based on messages sent by victims via mobile devices and on social media.

THE UNOFFICIAL FUTURE

THE OFFICIAL FUTURE

WHITE NOISE
THE FAILURE OF THE NEW DATA ANALYSIS METHODS

The purported revolution in our thinking is a flop. Wrongly programmed algorithms lead to a series of spectacular mishaps. The promise of more transparency remains unfulfilled. Since the increase in processing capacity is unable to keep pace with the growth of storage space, and since the storage space itself is not large enough to contain all the data being created, in relative terms our level of knowledge deteriorates. We can only look on helplessly as the sea of data moves inexorably towards us. Moreover, the analysis of ever-growing numbers of data sets does not necessarily lead to more precise results; because of so-called "overfitting" it can even lead to worse predictions. The availability of endless amounts of inaccurate data lends "statistical" credence to incorrect statements – thus leading us further away from the goal of more objectivity. And with an increased awareness of how data is collected and used; more and more people are refusing to allow access to their private data or are changing their behaviour to skew results.

THE AGE OF TRANSPARENCY
THE DATA WORLD KEEPS ITS PROMISES

A hitherto unseen clarity is starting to prevail. Our improved and data-driven knowledge of many things that are happening in the world makes for an unprecedented level of objectivity in the decision-making process. Personal effort is also starting to pay off more and more, as it is being appreciated. Everything is getting easier – from shopping to planning what to do with your life. For the future is predictable, since the consequences of our actions can be calculated in real time. Some decisions about the ideal course of action can be left to computers. This means we can use our minds for more important questions. In addition, completely new sectors of the economy are created as it becomes possible to offer customers ultra-personalised products. The world around us is equipped with sensors that can react intelligently to our needs and wishes, recognising them and almost always fulfilling them instantly while at the same time providing for a crime-free society. Moreover, the new and much more intelligent use of resources brings about a dramatic increase in efficiency and thereby solves a whole range of environmental problems.

IV: FIELDS OF ACTION FOR SURVIVING IN THE DATA WORLD: THINK FOR YOURSELF, CUT DOWN ON COMPLEXITY, MAKE CONSCIOUS USE OF ALGORITHMS!

HAVE FAITH IN YOUR OWN MIND. GAIN EXPERIENCE AND MAKE YOUR OWN DECISIONS.
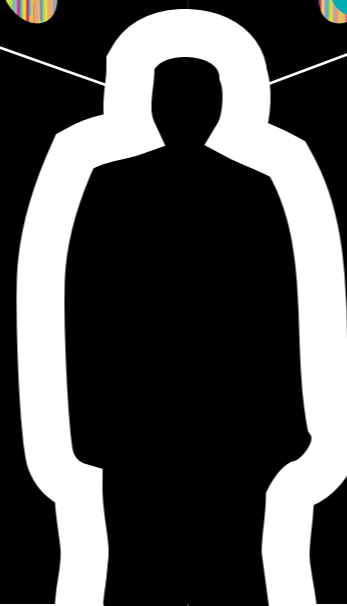
USE ALGORITHMS ONLY WHEN THEY OFFER ADDED VALUE. CHECK FACTS CRITICALLY.

MAKE ALGORITHMS TRANSPARENT.

PROMOTE THE SKILL OF MINIMISATION. TAKE MORE BREAKS.

Even as data processing helps us to learn more about the world and its social and technical systems, the sense of an increasing lack of transparency prevails – reinforced by the knowledge that there is a constantly growing volume of unprocessed data that we know absolutely nothing about. However, this in itself could prove to be useful, for it forces us to continue to think for ourselves instead of simply passing on decisions to a technical environment that is not as intelligent as we thought it was: the captain takes the wheel himself in the fog. Moreover, complete transparency would create new social issues. Our society is based on a series of covert and coded social interactions. If these were to become fully visible, society would have to be completely reorganised. For if we no longer had the privacy to have a good moan about our neighbours, this would mean the loss of valuable outlets for our emotions. A further danger in a society driven by algorithms would be of an increasing standardisation and a loss of variety. In biological systems it is that very variety that has invariably acted as a stabiliser and a safeguard against risks, while in business and politics it has always been the bedrock of innovation.

DIRECT INNOVATION TOWARDS REDUCING COMPLEXITY: LESS IS MORE.

BECOME AWARE OF THE MARKET VALUE OF PERSONAL DATA AND MANAGE IT ACCORDINGLY.

MAKE USE OF THE ADVANTAGES OF THE DATA WORLD IN SELECT AREAS AND PLUNGE HEADLONG INTO THE SEA OF DATA IN THOSE AREAS.

FAVOUR SIMPLE TECHNOLOGY.