KNOWLEDGE.ESSEC.EDU

ESSEC KNOWLEDGE REVIEW

JANUARY 2021

Artificial intelligence and the technological transformation

EDITORIAL

"By far, the greatest danger of Artificial Intelligence is that people conclude too early that they understand it."

- Eliezer Yudkowsky, founder of the Machine Learning Research Institute

The mention of artificial intelligence might conjure up images of robots, smart houses, or job automation. It might spark interest, excitement, or even trepidation. Whatever your reaction, it is unlikely to leave you indifferent: Artificial Intelligence systems touch virtually everyone, and the world of tomorrow promises to be exciting.

Our world is evolving rapidly, and we have the opportunity – the responsibility – to be a part of the change. Digital technologies and Data Science are impacting how we interact, learn, and work. Modern problems require modern solutions, and these raise a lot of questions.

ESSEC's motto is "Enlighten. Lead. Change", and we believe that it is our responsibility to prepare future leaders for the challenges we face as a society – and also to see challenges as opportunities. While the world of tomorrow still holds many unknowns, we believe in the power of curiosity. This is why, with this special issue, we aim to share the knowledge of ESSEC professors and provide a better understanding of how artificial intelligence and the digital transformation will impact society.

This special issue of ESSEC Knowledge contributes to the new RISE strategy at ESSEC, which positions our school at the interface of science, management and societal change. One element of the RISE strategy is the creation of the *Metalab for Data, Technology & Society,* an interdisciplinary lab for scientific research and pedagogical innovation. With this, we invest in research studying the methodologies, uses and impacts of artificial intelligence, using an interdisciplinary, reflective and responsible approach.

Highlighting the wide-ranging influences of artificial intelligence systems, this special issue features articles from professors in analytics, decision sciences, econometrics, economics, entrepreneurship, finance, information systems, management, marketing, philosophy, and statistics. ESSEC professors have provided their expert analyses on topics ranging from the use of artificial intelligence in human resources and marketing, data ownership, the impact of AI on businesses and decision-making, and the ethical dilemmas and potential discriminations posed by AI-powered governance and decision support systems. We discuss how AI and new technologies may impact the world of work, from HR, analytics, and marketing, to information governance, to how research is conducted, to their uses in sustainability initiatives. We also cover how it could impact our daily lives, discussing how it could be used in credit decisions, the interplay between AI and gender, and other societal implications, including the issue of misinformation spreading on social networks. Together, these provide a peek into what tomorrow could look like as we unleash the power of technology and algorithms - and also learn how to use this power for good.

The articles included in this special issue provide insight into our world's scientific transformation, stimulating discussions and debates so that we imagine what the next chapter of humanity could look like. In the words of the late Okwui Enwezor, who recently curated the Venice Biennale A*ll the World's Futures,* we ask "how can the current disquiet of our time be properly grasped, made comprehensible, examined, and articulated? [...] How can [we] make sense of the current upheaval?"

Julia Smith, Editor-in-Chief of ESSEC Knowledge
Professors Guillaume Chevillon and Julien Malaurent, Academic Co-Directors of the Metalab for Data, Technology & Society

TABLE OF CONTENTS

THE ARTIFICIAL INTELLIGENCE REVOLUTION IN ANALYTICS, AND WHAT IT MEANS FOR BUSINESS
Nicolas Prat,
Department of Information Systems, Decision Sciences and Statistics

BUILDING YOUR HR ANALYTICS STRATEGY: HOW TO SUCCEED AND WHAT TO WATCH OUT FOR
Valery Yakubovich,
Department of Management

MARKETING AND ARTIFICIAL INTELLIGENCE: PITFALLS AND POSSIBILITIES
Arnaud de Bruyn,
Department of Marketing

SUSTAINABLE DEVELOPMENT THANKS TO THE DATA FOOTPRINT
Jeroen Rombouts,
Department of Information Systems, Decision Sciences and Statistics

DECISION-MAKING IN THE ERA OF AI
Ivana Ljubic,
Department of Information Systems, Decision Sciences and Statistics

WHO SHOULD OWN YOUR DATA? AN ECONOMIC ANALYSIS
Wilfried Sand-Zantman,
Department of Economics

28

DO WE STILL NEED RESEARCHERS?

Vivianna Fang He,
Department of Management

32

ARTIFICIAL INTELLIGENCE AND INFORMATION GOVERNANCE: A PROMISING PAIR?

Isabelle Comyn-Wattiau,
Department of Information Systems, Decision Sciences and Statistics

36

AI FOR PERSONAL DEFAULT PREDICTION: PROMISES AND CHALLENGES AHEAD

Andras Fulop and Laurence Daures-Lescourret,
Department of Finance

40

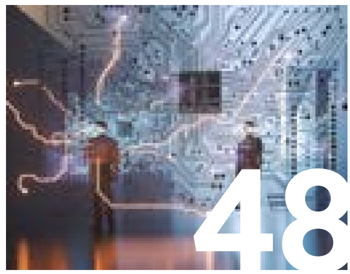
THE LINK BETWEEN DATA MATURITY, DATA TALENTS, AND RENEWABLE ENERGY

Jeroen Rombouts,
Department of Information Systems, Decision Sciences and Statistics

44

THE ETHICS OF ARTIFICIAL INTELLIGENCE

Laurent Bibard,
Department of Management



48

GENDER INEQUALITY: IS AI A BLESSING OR A CURSE?

Estefania Santacreu-Vasut,
Department of Economics

52

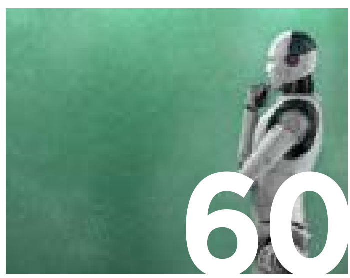
THE ETHICAL DILEMMAS OF AI-POWERED GOVERNANCE SYSTEMS

Julien Malaurent and Thomas Huber,
Department of Information Systems, Decision Sciences and Statistics

56

TOWARDS A POLICY OF ALGORITHM SECURITY?

Guillaume Chevillon,
Department of Information Systems, Decision Sciences and Statistics



60

SO MANY WAYS TO BE WRONG ABOUT THE FUTURE OF AI, BUT IT'LL OUTSMART US IN THE END

Fabrice Cavarretta,
Department of Management

THE ARTIFICIAL INTELLIGENCE REVOLUTION IN ANALYTICS, AND WHAT IT MEANS FOR BUSINESS

__Nicolas Prat__ is Associate Professor of Information Systems and Information Technology at ESSEC Business School. He holds a PhD in information systems and an accreditation to supervise research from Paris-Dauphine University, as well as a Master of Science in management from ESSEC Business School. His expertise areas include business intelligence and business analytics, design science research, and data traceability and provenance. He has published numerous papers on this topic, including in the Journal of Management Information Systems, Decision Support Systems, and Expert Systems with Applications. He is a member of the editorial board of the Journal of Database Management. He has been a visiting scholar at the University of Alicante (Spain) and at the MIT Sloan School of Management.

The dramatic progress of artificial intelligence (AI) is affecting many sectors and functions of business. In his research paper "Augmented Analytics", Professor Nicolas Prat focuses on the impact of AI on analytics. How is AI revolutionizing analytics, and what are the opportunities and challenges for managers?

"Analytics" refers to the technologies and processes for collecting, blending, modeling, analyzing and visualizing data in order to gain insights and make better decisions. Different types of analytics have been defined: descriptive, diagnostic (inquisitive), predictive, and prescriptive.

The term "analytics" became popular in the mid-2000s. In the early 2010s, the big data revolution gave rise to big data analytics. Self-service business intelligence (BI) empowered business users to analyze their data and generate their own visualizations without systematically resorting to the IT department. Today, AI (more specifically, machine learning and natural language processing) is bringing about a new revolution in analytics. Gartner calls this revolution "augmented analytics". Others talk about "the cognitive generation of decision support", "smart analytics" or, more simply, "AI-powered analytics". Predictive analytics (and, more generally, advanced analytics) has traditionally relied on machine-learning algorithms, like neural networks, for the development of models. However, what is new with AI-powered analytics is the scope of applications of AI throughout the analytics cycle. This includes, for example, the application of machine learning to find the best machine-learning model (applying machine learning to the automation of machine learning...).

Artificial Intelligence Is Permeating the Whole Analytics Cycle

The typical analytics cycle is composed of the following phases:

- Phase 1: identification of the business problem addressed by analytics, as well as the opportunities of big data analytics for the business.
- Phase 2: data preparation ("wrangling"), decomposed into data profiling (assessment of data quality) and transformation.
- Phase 3: data analysis. In this phase, a distinction is made between data discovery (that may be performed

by analysts or decision makers using self-service BI tools) and modeling (model building and evaluation by data scientists).
- Phase 4: deployment of the models in production systems.
- Phase 5: decision based on the insights from the data.
- Phase 6: action implementing the decision.
- Phase 7: monitoring of the results of actions, e.g. to improve the models.

The examples below give a sample of applications of AI to the analytics cycle. Although these applications differ in maturity, they are already realities. Many more are likely to emerge as AI continues to approach human intelligence.

In phase 2, some tools guide users in data preparation, e.g. by suggesting data transformations to clean or standardize the data. AI-enabled automation here may significantly improve efficiency, as data preparation often takes more than two-thirds of the time in the analytics process.

In phase 3, machine learning and natural language processing are significantly transforming data discovery. Examples include the enhancement of visualizations with advanced analytics (clustering, forecasting), guidance in data discovery (suggestions of relevant visualizations based on the current stage of data discovery), natural language data querying, and natural language generation (automatic generation of the text summarizing the insights from a visualization). In modeling, some tools use machine learning to automate the generation and assessment of machine-learning models, traditionally a major role of data scientists.

In phases 5 and 6, some operational decisions are already made and implemented automatically thanks to AI. High-frequency trading is a case in point.

Democratization of Analytics

Decision support systems (DSS) appeared in the 1970s. They have evolved to give business users more and more freedom in analyzing their data to gain insights and make relevant and timely decisions. Executive Information Systems (EIS) were the early version of electronic dashboards, aimed at top managers. Thanks to the development of data warehouses, business users were able to generate reports independently, i.e. without systematically resorting to the IT department. With self-service BI, the democratization of analytics was pushed a step further, with tools like Tableau, Qlik, or Power BI, facilitating data discovery and visualization. However, these tools were still limited to traditional BI tasks: reporting, OnLine Analytical Processing (OLAP), and dashboards. “Smart data discovery” (a term coined by Gartner) enhanced self-service BI tools with advanced analytics. For example, IBM Watson Analytics heavily relies on predictive analytics. Smart data discovery lowered the barrier to entry to advanced analytics for domain specialists, business users, and managers, virtually making them into what Gartner calls “citizen data scientists”. With the last generation of AI-powered analytics, the democratization of analytics extends beyond data discovery. Thus, at least in principle, you do not need to be an IT professional to clean data or a data scientist to generate and evaluate machine-learning models.

The claim that data-driven decision making leads to better decisions and that mastering analytics leads to better performance is now commonly accepted and supported by empirical evidence. In the age of big data and the Internet of Things, data relevant for decision making are more and more abundant, more and more diverse, and decisions should be made more and more quickly. This context, coupled with a lack of data scientists, requires business users to play an increasing role in the analytics process. AI-powered analytics makes it possible. However, the rise of AI-powered analytics raises several challenges. Governance of analytics and, in particular, ensuring data quality, is a major challenge.

Data Quality and Governance

The quality of analytics is heavily dependent on the quality of input data ("garbage in, garbage out"). In practice, data are often incomplete, inaccurate, inconsistent, or biased. The fact that AI-powered analytics democratizes data preparation multiplies the risks of bad data. Moreover, data quality influences trust in AI-powered analytics, as it influences trust in AI more generally. Humans still have a major role in ensuring data quality, e.g. in the interpretation of "bad data".

Beyond data quality, AI-powered analytics makes the governance of the whole analytics cycle more crucial and challenging. This cycle is complex, may be composed of several sub-cycles, and may be instantiated in many different ways. It involves many stakeholders and tools. The democratized access to analytics, redefining the roles of stakeholders, further complicates the orchestration of the analytics cycle.

With the digital revolution, some tasks traditionally assigned to the IT department have moved to other functions. However, this department has a crucial role to play in the governance of IT in general, and analytics in particular, e.g. through the definition of processes, roles, and common standards, in dialogue with managers.

Where Artificial Intelligence (Still) Cannot Beat Humans

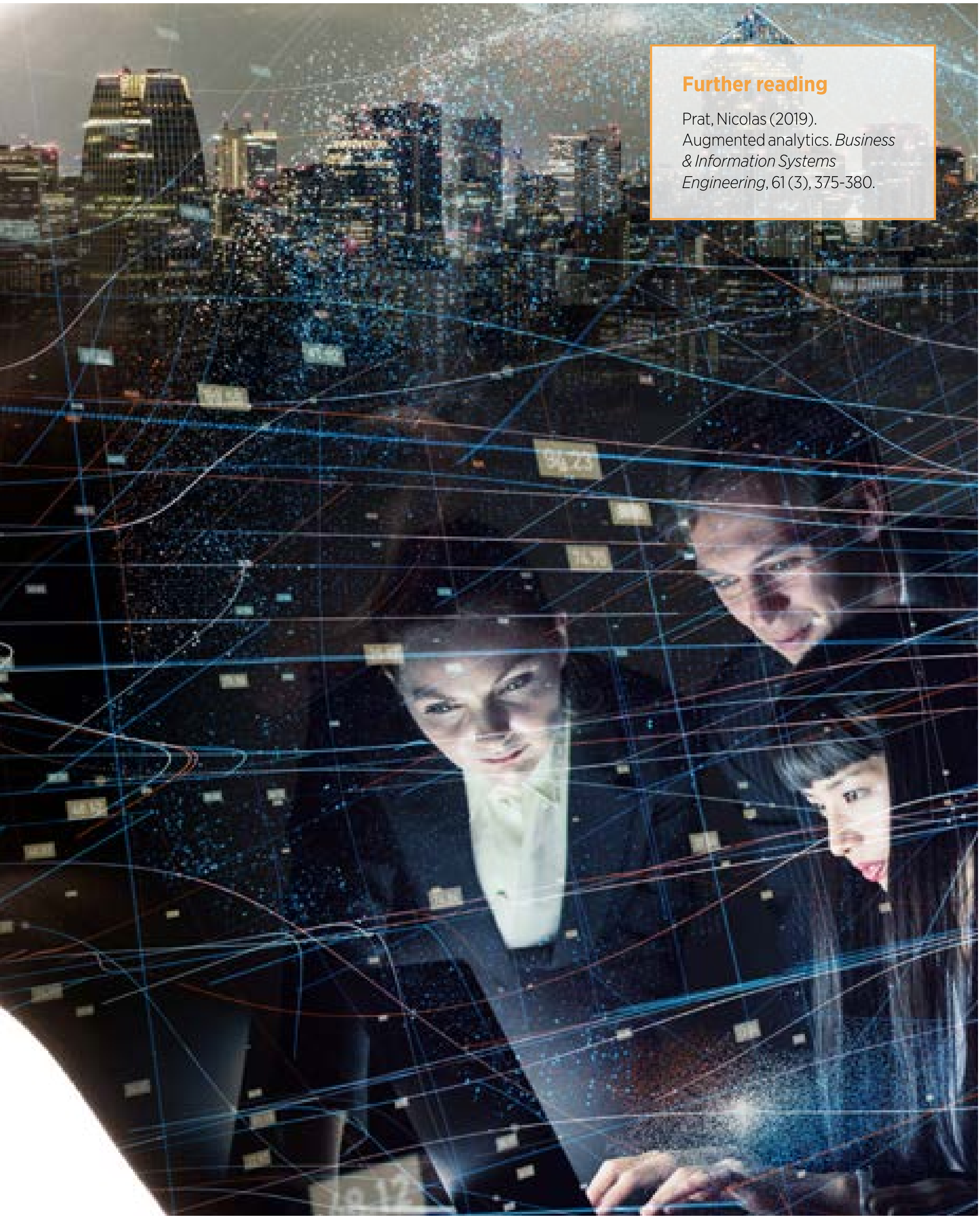
By enabling the automation of tasks traditionally accomplished by human agents, the advent of AI-powered analytics brings about a redefinition of roles. The traditional roles in the analytics cycle should evolve by considering the skills that still give humans a competitive advantage over machines. Even if the accelerating speed of technological progress is gradually pushing the limits of AI, humans typically outperform machines in the following skills: question asking (finding problems as opposed to solving them), creativity, interpersonal skills, and social intelligenace. These skills suggest ways of evolving the roles of data scientists, analysts or business users in the analytics process. Data scientists, threatened by AI in their very function of model creation and evaluation, could strengthen their role in the early phase of the analytics cycle: identifying the business questions that the machine-learning models will help answering. Data scientists, as well as analysts, should sharpen their storytelling skills. Like the art of writing novels, storytelling with data requires creativity. Finally, business users could exploit their social skills to make analytics a more collective process. Contrary to the group decision support systems (GDSS) of yesteryears, the main analytics systems of today very much focus on individuals. Using virtual or augmented reality (immersive analytics), business users may analyze data collectively and discuss insights before making a common decision.

If properly managed, AI-powered analytics provides an opportunity to analyze more data, more efficiently, and, ultimately, to make better decisions. Beyond the challenges mentioned above, many questions require further investigation by researchers and practitioners, such as:

- Should all business users take the role of "citizen data scientist", or should a specific category of business users play this role?
- To what extent does AI improve the perceived usefulness and perceived ease of use of analytics systems?

- What are the key determinants of success of AI-powered analytics?
- What are the main factors affecting the credibility of the insights that AI-powered analytics generates and trust in the decisions that it suggests or makes?

Originally published on October 18th, 2018



© GettyImages - Metamorworks

Further reading

Prat, Nicolas (2019). Augmented analytics. *Business & Information Systems Engineering*, 61 (3), 375-380.

BUILDING YOUR HR ANALYTICS STRATEGY: HOW TO SUCCEED AND WHAT TO WATCH OUT FOR

Valery Yakubovich *is Professor of Management at ESSEC Business School and Senior Fellow at the Wharton Center for Human Resources. The results of his research have been published in* American Sociological Review, Human Relations, Organization Science, California Management Review, *Harvard Business Review, and other journals and edited volumes. His current projects explore innovative organizational and management practices in the context of digital transformation, and coevolution of careers and social networks in organizations.*

The discussion surrounding the digital transformation in management has moved from big data to machine learning to artificial intelligence at an astounding speed. Yet the gap between promise and reality remains wide: 41% of CEOs report that they are not at all prepared to use new data analytics tools, and only 4% say they are "to a large extent" prepared (1). In their recent article, ESSEC management professor Valery Yakubovich and his colleagues Peter Cappelli and Prasanna Tambe from the Wharton School at the University of Pennsylvania identify four challenges in using artificial intelligence techniques in human resources management and propose practical responses to these challenges.

All That Glitters Is Not Gold

AI's increasing sophistication when it comes to predictive analytics makes it very interesting for human resources. It could be applied to a variety of HR practices, like recruitment and selection, training, and employee retention. But since HR involves dealing with people, the questions are important and nuanced and don't have cut-and-dry answers. Further complicating matters is the fact that HR datasets tend to be much smaller than those in other domains, such as market research, and data science techniques perform poorly when predicting relatively rare outcomes. Firing someone for poor performance is one example of an outcome that happens relatively rarely in companies, but one that has important implications for individuals and society.

The problems that AI faces in its HR applications tend to fall into four main groups: the complexity of HR problems, small datasets, ethical and legal considerations, and employee reactions. We explain how these apply at each stage of the AI life cycle, from data generation, to machine learning, to decision-making, and include questions to ask yourself when designing an AI strategy for HR management.

The life cycle of an AI project

1. Generating Data

Gathering the data for your AI algorithm can be complicated. Take the seemingly straightforward question, "What constitutes a good

© GettyImages - Orbon Alija

employee?", a question that becomes less straightforward when you dig a little deeper. Job requirements can be broad and hard to specify for an algorithm. There is also the question of bias: an AI algorithm might be able to identify relationships between employee attributes and job performance, but if, for example, a company has historically hired and promoted white men, the algorithm could predict that white men will be the highest performers and inadvertently discriminate against other candidates, even if those candidates are highly qualified. Measuring performance can also present challenges: Who assesses performance? What is it based on? We work in an interconnected ecosystem, so performance is also impacted by factors like our colleagues, job resources, and company culture. Ideally an algorithm would include multiple indicators of performance, but creating an aggregate variable to represent performance is difficult. Therefore, do not seek perfect measures as they do not exist, but choose instead reasonable ones and stick with them.

There is also some selection bias in assessing employees, as often only those that were hired are included in the dataset. Most companies do not keep records of all of the data that they accumulate. To build a larger dataset, aggregate information from multiple sources and over time, including from candidates who are screened out.

Before launching a new digital HR project, determine the necessary and available data that can be extracted and transferred into a usable format at a reasonable cost. Sharing data across units must become a priority in the short-term; to evaluate employees' performance, you must incorporate the company's business and financial data. Invest in data standardization and platform integration across your company in the long run.

Do you have enough data to build an algorithm? Small datasets are often sufficient for identifying causal relationships, which managers need to understand in order to act on insights. Therefore, the less data you have, the more theory you will need (drawing from management literature, expert knowledge, and managerial experience). Randomized experiments are not to be neglected in order to test causal assumptions.

If other companies are making their data available for machine learning, make sure that your context is not too distinct so that the algorithm built on data from elsewhere will be effective for your own organization. You can also use social media as an alternative source of data: some employers use it for hiring, others to identify problems such as harassment. HR stakeholders also must take privacy considerations into account and see under what conditions employee data can be used.

2. Using Machine Learning

Consider the example of using machine learning in the hiring process: we might look at which applicant characteristics have been linked to stronger performance in the past and use this to inform our hiring decisions. Using a machine learning algorithm might end up working better than conventional strategies but it poses a problem of self-selection: the ability of the model to "keep learning" and adapt to new information disappears when the flow of new hires is constrained by the predictions of the current algorithm. To address this problem, it could be useful to periodically retrain the algorithm using data on the performance of candidates that do not fit its criteria.

Another possible issue is that using algorithms in selection could reduce the range on the variables of interest, potentially masking true relationships. For example, if a hiring manager makes its decision based on university grades, they might then have a hard time finding a link between grades and performance, for the simple reason that there is little variability in employees' grades and so the relationship is not as clear.

There are also potential ethical issues with the use of algorithms in HR decisions. For example, if we consider the difference between majority populations and minority populations, algorithms that maximize predictive success for the population as a whole may be less accurate in predicting success for the minority population. Generating separate algorithms for both might lead to better outcomes, but also to conflicts with legal norms of disparate treatment. Thus, the effective implementation of machine-learning algorithms requires a review of labor laws.

3. Decision-Making

When choosing between two candidates that are both qualified for the position, the hiring manager has to make a tough decision. Suppose an algorithm determines that one candidate is an 80% match for the position and the other one is a 90% match. Is a 10% difference large or small, taking into account some very likely measurement errors and biases? In order to mitigate some of these issues, we could introduce random variation, which has been an unrecognized but important mechanism in management. Contrary to popular belief, research shows that employees perceive random processes as fair in determining complex and thus uncertain outcomes. Therefore, if both candidates are strong, it makes more sense to make a random choice. **In other words, randomization should be an AI-management tool.**

Employee buy-in is also a key part of the equation, as they will be impacted by changes in the decision-making process. How will employees react to decisions made by an algorithm instead of a supervisor? Even if employees are not always committed to the organization, they might be committed to their manager. Let us

© GettyImages - Artystarty

look at the following example. In the workplace, if your supervisor assigns you to work on the weekend, you might do it without complaining if you think your supervisor is generally fair. When the work schedule is generated by a program, you might react differently, as you don't have a preexisting relationship with the algorithm. That being said, some decisions are easier to accept from an algorithm especially when those decisions have negative consequences for us, such as increased prices, as the decision feels less personal.

So where do we go from here?

These are a few questions you should ask yourself before using AI technologies in HR management. In sum, remember:

1. Causal explanations are essential for analytics and decision-making in HR because they can ensure fairness, be understood by stakeholders, and are ethically and morally defensible.

2. Companies have to accept HR algorithms' relatively low predictive power.

3. Randomization can help with establishing causality and partially compensate for algorithms' low predictive power.

4. Formalizing algorithm development processes and involving stakeholders in the process will help employees form a consensus about the use of algorithms and accept their outcomes. ■

Originally published on October 29th, 2018; updated in December 2020

Further reading

Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, 61 (4), 15-42.

MARKETING AND ARTIFICIAL INTELLIGENCE: PITFALLS AND POSSIBILITIES

Arnaud de Bruyn *is Professor of Marketing. Arnaud's research focuses on marketing analytics, direct and interactive marketing, and marketing engineering. He has been published in* Marketing Science, Management Science, and Information Systems Research, *amongst other journals. He has multiple scientific prizes for his work on viral marketing. Arnaud is a Distinguished Visiting Research Scholar at the Institute for the Study of Business Markets. He has extensive experience in the non-profit sector, where he has assisted over 40 non-profit organizations to optimize their direct marketing. He holds a PhD in Business Administration from Penn State University.*

Written by Julia Smith, Editor-in-Chief

The rise of artificial intelligence has naturally seen people applying it (or attempting to apply it) in countless ways, with varying degrees of success. While artificial intelligence can be a powerful tool- in the right hands, in the right situation- it is not as easy as implementing a system and tapping a few keys. This is perhaps especially true for environments that deal with human behavior, like marketing. As the saying goes, "with great power, comes great responsibility": and marketing managers must be aware of its potential pitfalls to avoid problems. Equally important is the need to know how to properly deploy their AI tools to avoid squandering both its potential and their company efforts and resources. By understanding AI's pitfalls, marketing managers can make the most of its opportunities.

Arnaud de Bruyn and his colleagues Vijay Viswanathan (Northwestern University) Yean Shan Beh (The University of Auckland; Xiamen University Malaysia), Jürgen Kai-Uwe Brock (Fujitsu Global), and Florian von Wangenheim (ETH Zurich) explored this in their recent article.

So far, AI's biggest advancements in the business world have been related to deep learning, referring to complex, multilayered (i.e., deep) neural networks, solving difficult problems with predictive analytics. The more layers in a neural network, the more complex it is, and more "layered" networks can identify and learn more complex relationships between variables. This means artificial intelligence can learn to uncover relationships that existing statistical techniques cannot detect and that it can learn to do so autonomously. This is the main selling point of contemporary AI algorithms.

While the ability of AI algorithms to autonomously create models is its strength, it is not without its challenges when it comes to putting it in action. These challenges are: a lack of common sense, objective functions, a safe and realistic learning environment, biased algorithms, understandable and controllable artificial intelligence, the paradox of automation, and knowledge transfer.

© GettyImages - Metamorworks

Lack of common sense

What do we mean by a lack of common sense? It is not an insult to its programmers or to those operating it; no, we mean that the algorithm itself lacks what we humans call "common sense". We know that emotional intelligence is important, and indeed AI systems are increasingly able to recognize people's emotions, through image recognition, voice analysis, or text analysis. But recognizing emotions is a far cry from understanding and feeling them. An AI system could learn that the words "queen" and "crown" are linked, and could even use them appropriately in a sentence, but the meaning of the words and sentences would be lost on it. Anything they have approaching common sense must be programmed into them by a person, which becomes a problem when it comes to objective functions.

Objective functions

An objective function is one that specifies the result that the AI algorithm aims to optimize (Sutton and Barto, 2018). In a marketing context, this could look like profit maximization or customer retention. The "freedom" of AI from common sense hinders its ability to define an objective function. It might be that humans understand something implicitly, but then have a hard time translating this for the algorithm. This might go awry: an autonomous car directed to "get to the airport ASAP!" might get there in record time, but having mowed down pedestrians and sped through red lights on its way. While the previous example is obviously extreme, we have already seen consequences of this play out in real life, with gender- or racially-biased systems. An outcome like profit maximization cannot be considered without allowing for the legal, moral, and ethical implications, which marketing stakeholders need to keep in mind when building and implementing their systems.

Safe and realistic learning environment

As you can imagine, all this is easier said than done. Knowledge transfer from the expert to the algorithm and vice versa is one of the biggest problems facing AI today, and the potential for costly mistakes is enormous. To avoid the fallout, it is important for AI algorithms to learn in a safe, realistic environment. Safe, in that if they do make mistakes, there is less impact on the business, and they avoid the marketing equivalent of running a red light. Realistic, in that the data resembles what they would receive in a real-life situation. This presents a challenge in marketing, because customers can be unpredictable, and a new factor (like, say, COVID-19) can throw a wrench into the best-laid marketing campaigns. While it might be tempting to think that AI reduces or even eliminates our need to understand customer behavior, it is the opposite: we need detailed customer behavior theory more than ever, as this will help us better configure our AI algorithms.

Biased algorithms

This brings us to another limitation to AI's use in marketing: its potential to be biased. Of course, the algorithm itself is not prejudiced, but if it is powerful enough, it could identify a characteristic like race or gender on its own and make biased predictions. How so? It might pick up on other information that acts as a proxy to the factor in question, like education or income, thereby unintentionally replicating the biases

that are found in the data. In a marketing context, this could lead to outcomes like a price-optimization algorithm that aims to charge women more or an advertising algorithm that targets a vulnerable population. This has legal implications as well as the obvious ethical ones. Complicating the problem is the fact that adding the sociodemographic variable in question to the model in an attempt to clarify it could just make it easier for the algorithm to make prejudiced predictions. If marketing stakeholders do not properly understand the algorithms they are using, they might not know to challenge these troubling predictions.

Understandable artificial intelligence

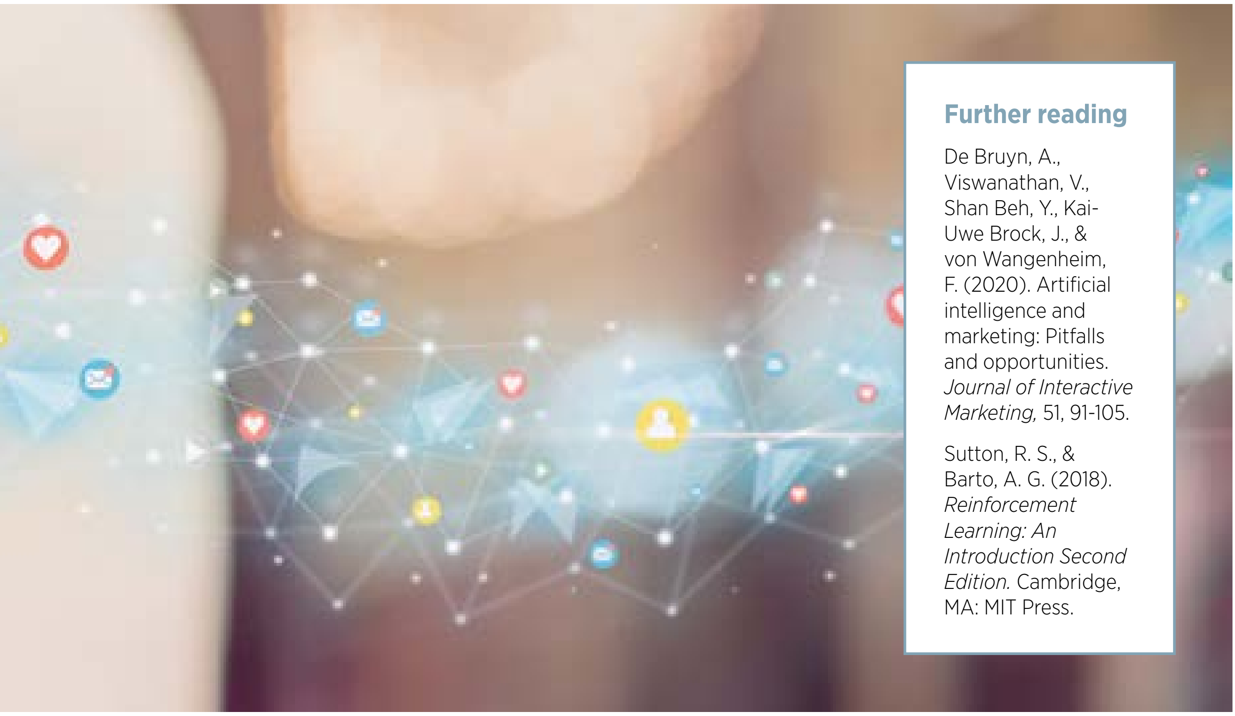
The ability to understand and explain the model is another factor in the uptake of AI. If you are going to use an AI model, you need to understand why it makes the predictions it does, and to be able to interpret what the model is doing. More specifically, an AI's human "handlers" need to be able to explain: 1) the purpose of the model, 2) the data it is using, and 3) how the inputs relate to the outputs. By understanding this, it is also possible to know why the AI system is preferable to a non-AI system.

Controllable artificial intelligence

Using the term "handlers" above was intentional: an AI system must be able to be controlled and overridden. This might conjure up images of I, Robot and killer robots, and while the reality is rather less lethal, it is still serious. One recent example is that Uber's pricing algorithm responded to the crush of people fleeing the scene of the June 2017 terrorist attack in London by adapting (read: increasing) the ride prices to more than double the typical fare. Anyone who has taken Uber is unfortunately familiar with their surge pricing system, but in the aftermath of a terrorist attack, it made Uber seem like ruthless profiteers. However, Uber's monitoring system quickly flagged the problem, and they had mechanisms established that allowed them to override the algorithm within minutes. They were also quick to communicate about what was going on, made rides free in that area, and reimburse those affected. Alas, the damage was done. This situation left a black mark on their reputation and serves as a warning to marketing managers that any algorithm they implement needs to be constantly monitored and have the possibility to be overridden built in.

The Paradox of Automation

The purpose of automation is to replace the role of humans, aiming to make tasks faster and more accurate and leaving people free to do more complex work. The downside to this is that then people don't have experience with those simpler tasks and don't have the opportunity to gradually build up their



© GettyImages - Sitthiphong

Further reading

De Bruyn, A., Viswanathan, V., Shan Beh, Y., Kai-Uwe Brock, J., & von Wangenheim, F. (2020). Artificial intelligence and marketing: Pitfalls and opportunities. *Journal of Interactive Marketing,* 51, 91-105.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction Second Edition.* Cambridge, MA: MIT Press.

expertise and skills. In marketing, this could mean that those in marketing, from customer service agents to market research analysts, miss the opportunity to hone their skills on simpler and more repetitive tasks that allow them to better understand customers and their needs, and are left dealing with only the most complicated and unique cases. It remains to be seen what implications this would have for the quality of service and work.

The next frontier of AI and marketing: transferring and creating knowledge

What sets AI apart from traditional statistics is its ability to execute higher-order learning, like uncovering relationships between indicators to predict the likelihood that an Internet user will click on an ad, and to do so autonomously. Being able to create knowledge like this is a huge advantage of AI. However, the transfer of knowledge from the AI model to the expert and vice versa is a major weakness of AI. Since marketing deals with human behavior, this requires a lot of common sense, which, as we now know, is not the forte of AI models. Since this kind of knowledge is often more implicit, dealing with social codes and norms, it is also harder to program into an AI model. The machine will also be able to pick up on links that it needs to transfer back to the human expert, especially so that the experts can identify flaws in the system and understand how it is operating. An AI system that is able to create and transfer knowledge back to the human expert is thus the Holy Grail of AI technology.

Takeaways

So what is a marketing manager who wants to use AI to do? There are a few key points to keep in mind:

1. Understand the purpose of implementing the AI system. What are you aiming to accomplish?
2. Identify the added value of the AI system. What does it add over and above human capabilities?
3. Understand what your AI system is doing. What data is it analyzing? How is it producing the results?
4. Examine the system for bias. Does your system have any built-in biases?
5. Communicate: ensure that relevant stakeholders (consumers, employees) have the possibility to observe and interact with the AI system, to build trust, ensure reciprocal knowledge transfer, and practice.

SUSTAINABLE DEVELOPMENT THANKS TO THE DATA FOOTPRINT

Jeroen Rombouts *has been a professor at ESSEC Business School since 2013. He holds the Accenture Strategic Business Analytics chair. He obtained his PhD in econometrics in 2004. As a full professor of statistics and econometrics, he teaches big data analytics and data science courses at the MBA and Master's programs in Paris and Singapore. He researches predictive analytics, in particular time series analysis and forecasting. He has published numerous scientific articles, frequently organizes workshops, and is associate editor of several quantitative scientific journals. His research is funded by the European Commission and the ANR. He also consults on strategic business analytics and data science topics. Prior to joining ESSEC, Jeroen was an associate professor at HEC Montreal.*

Co-author information

Reda Gomery: *Reda Gomery is the vice president of Data at A.T. Kearney, a global management consulting firm. Contact: Reda.Gomery@kearney.com*

Since the start of the COVID-19 pandemic, companies have had to accelerate their digital transformation. This implies increased investments, so substantial that they require C-level support. The stakes are high for organizations. From accelerating sales to optimizing operational processes, digital impacts the value chain in every aspect. If the digital revolution generates an inevitable modernization of companies and a hope of value generation, it also provokes a major challenge for organizations: Data.

Data from transactions, customers, products, etc. invades the daily operations of organizations, constituting a potentially valuable asset, but above all an important challenge in terms of governance and management. Organizations must increase the understanding of these data as part of their transformation.

In the very short term and in an uncertain time, data becomes more crucial than ever to identify the levers of performance of companies. Optimizing costs, increasing business revenues, and driving process efficiency are all initiatives based on the availability of relevant data. As the decision cycles accelerate, many decision-makers will no longer be able to drive their businesses with approximate and often inaccurate data. Having good data - and just in time - has become a pressing necessity. But this prospect seems attainable only if the data heritage is better mastered. This is precisely the purpose of the "Data Footprint" method designed by Kearney and Essec. Evaluating the data footprint now constitutes an essential approach to secure investments and increase control over data assets.

The Data Footprint approach introduces a virtuous practice that aims to understand the data heritage, risks, challenges and limits linked to data within organizations. The Data Footprint is an evaluation process based on a 360° analysis of the data required as part of a company initiative steered by the entity in charge of Data Governance.

The aim of the Data Footprint is to assess the data assets to establish a risk assessment score. Based on multiple dimensions of analysis such as data quality or security, our method allows a quantified assessment of the data heritage in an organization. Today, the data heritage is still poorly controlled and exploited in many companies.

© GettyImages - Enot-poloskun

What is the quality level of critical data sets in the organization (e.g customers/suppliers' data)? What is the level of risk associated? What is the degree of control and ownership of data in the organization? These questions are often asked by decision makers without concrete answers based on a structured assessment. The complexity of information systems combined with the lack of governance make the data equation often complex and costly.

The Data Footprint allows companies to get a tangible data assessment across multiple dimensions in order to establish a risk score. The purpose of such a measure is to be able to accurately assess areas of weakness and to monitor data heritage improvements. The approach also allows internal and external benchmarks based on a standardized analysis grid.

The strategy for implementing a Data Footprint should be progressive while focusing on the critical data sets in the context of companies' major programs, projects or business transformation initiatives.

The approach should involve several collaborators, at least representatives of business lines and IT, who jointly use a score sheet based on the following five dimensions: accessibility and availability, quality, ownership, risks, and identification of the future users. The overall score calculated on these five dimensions can range between 0 and 15, the lower the score the higher the risk related to the enterprise initiative.

Consider as an example a company specializing in the distribution of electronic equipment to the general public through its distribution network of more than 2,000 stores. As part of its data strategy, the company decides to launch a priority project that deploys a "Customer-centric" approach in order to increase customer value. The objective is to capture a better understanding of customer preferences in order to meet their expectations. The company anticipates a significant potential risk linked to data (availability, quality, etc.) and decides to launch a Data Footprint approach.

The total Data risk score for this company was less than 5 in the evaluation exercise. On the recommendation of the Chief Data Officer in agreement with the rest of the team, the decision to launch the project is postponed pending the implementation of a specific data related action plan. This approach allowed the company to apprehend a major risk related to data on this project. Indeed, a rapid launch of this project without prior assessment would have potentially led to failure with economic consequences (losses estimated at a few hundred thousand euros). The approach also made it possible to initiate collaborative work around the data over the entire duration of this assessment (one month), and thus avoiding internal misunderstandings about the responsibilities of the various stakeholders (Business lines, IT teams, etc.). Finally, a clear action plan could be drawn justifying the investment of technical and human resources to upgrade the information system. ■

For a more technical version of this article or further details on the Data Footprint, please contact:

Reda Gomery, Vice President, A.T. Kearney, Reda.Gomery@kearney.com
Jeroen Rombouts, Professor, Essec Business School, rombouts@essec.edu

DECISION-MAKING IN THE ERA OF AI

Ivana Ljubic *is Professor at ESSEC Business School, where she is the academic director of the ESSEC & Mannheim EMBA program. She teaches Decision Analytics, Optimal Decision Making, Operations Research, and Supply Chain Management. She has been a professor or visiting scholar at the University of Vienna, the Robert H. Smith School of Business, TU Dortmund, TU Berlin, and Dauphine University. Ivana teaches and researches the role of Business Analytics in decision-making in the era of Big Data and AI. She obtained a PhD in Computer Science from the Vienna University of Technology, and a habilitation in Operations Research from the University of Vienna. Ivana has published more than 50 articles in leading scientific journals in operations research and management science.*

The term AI has been around for nearly 70 years, but it is only nowadays that AI is profoundly affecting and changing the way we live, work, interact and play. AI is on an unstoppable journey toward revolutionizing and disrupting diverse industries, and business leaders are embracing this trend with enthusiasm. So, why haven't we seen the rise of AI much earlier and is AI here to stay this time? Or is AI yet another new technology riding the hype wave? How can AI help us become better decision-makers?

First of all, what is AI?

Many researchers agree with Larry Tesler (a computer scientist who invented copy-paste while working at Xerox and worked on human-machine interaction) that "AI is anything that hasn't been done yet!". This means that the definition of AI changes over time. As we get used to previous advances in technology and computer science, what is called AI today, will be regarded as mainstream tomorrow.

Nevertheless, many will agree that, contrary to many "AI winters" that we have experienced in the past, now it is different: the time is ripe and AI is here to stay. This is because of the three major forces that have converged to enable the "AI revolution" we are facing today:

1) Increased processing and storage capabilities: we have seen a one trillion-fold increase in computing power between 1956 and 2015 with the rapidly falling cost of technologies, a trend that is continuing to this day. Nintendo consoles in 1983 are just as powerful as the Apollo Guidance Computer that brought the first humans to the moon in 1969. The 2010 iPhone 4 is just as powerful as the fastest machine in the world in 1985, the Cray-2 supercomputer.

2) Cloud computing: data storage and computing capabilities are now outsourced to the "clouds". Today, concepts like Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and even Machine-Learning-as-a-Service (MLaaS) are available at affordable prices. They provide low barriers to entry and allow businesses to quickly adopt new IT technologies.

3) Connectivity and availability of real-time data: large swaths of varied and real-time data can be collected


© GettyImages - Metamorworks

instantaneously, thanks to mobile phones, sensors, the IoT and other devices linked to the internet or intranet networks. This enables real-time prediction, prescriptions, automation and coordination of many complex processes.

How can we improve decision making by leveraging (big) data analytics and AI?

You might be excited by the recent technological advances around robotics, self-driving cars, flying taxis, etc. You might be equally terrified of military applications of AI, or of the potential threats related to face recognition or racial profiling. For now, let us focus on a more practical question: how can we improve everyday business practices by leveraging (big) data analytics and AI? This is where AI with Business Analytics (AI+BA), also called "AI-powered analytics", comes into play.

Business Analytics is the exploration and exploitation of an organization's data, with an emphasis on 1) data analytics, 2) predictions, and 3) prescriptions and decision support. Companies that use BA are committed to making better, more informed and data-driven decisions. BA is applicable across many different sectors: retail, transportation, entertainment, health-care, and energy, to only name a few. BA is a multi-disciplinary approach that combines computer science, statistics, mathematics, decision making and optimization. It allows organizations to eradicate some typical cognitive decision biases. And, more importantly, with data-driven decision making, it becomes possible to "mute the HiPPOs" (highest-paid-person opinions) in the room.

In organizations, there are typically three levels of decision making:

- strategic (involving long-term, non-routine and complex decisions made by senior management),
- tactical (dealing with mid-term, less complex decisions made by middle management), and
- operational (consisting of day-to-day routine decisions).

AI+BA can be successfully employed across all these levels, but with a different level of automation.

Let us take Netflix as an example: did you know that Netflix tracks not only movies that you are watching, but also the summaries you are reading and how much time you spend surfing titles and watching the trailers? In other words, Netflix knows the content you like before you know yourself. It is widely known that Netflix employs recommendation systems and personalized queues. Recommendation systems are indeed a standard procedure for doing digital-media distribution today. Recommendations delivered to the end-users can be seen as operational decisions, fully automated by algorithms and machines, and as the user, you may feel overwhelmed or pleased with them. But this is just the "AI'' or "predictive" part of Netflix's AI+BA approach. What is more, Netflix uses BA not only to suggest what you should watch, but to decide which projects to fund. Hence, important strategic decisions, like capital budgeting, are made based on predictions derived from the huge amount of collected user data. That way, the company uses advanced mathematical modeling and algorithms to decide on the content to be produced (which movie genres or types of characters will be the most popular ones?), on the amount of budget devoted

to different productions, and on project scheduling (when and where to shoot, and with which available resources, and when to schedule the release dates?). How successful is this approach? The figures speak for themselves: customer retention rate is 91% (compared with Hulu's 64% and Amazon Prime's 75%), and the total number of users has grown from 20M in 2011 to 186M in 2020!

Nowadays, many operational and routine decisions can be fully automated. The company UPS employed the Operations Research tools (the "prescriptive" part of BA) to automatically identify the most efficient routes for UPS drivers. UPS integrated these tools into drivers' navigation systems and fully deployed the technology in the U.S. in 2016. The results were 10 million gallons of fuel saved within a single year - the equivalent of saving 100,000 metric tons in carbon emissions!

The Paris-based company, Decision Brain, uses AI+BA to optimize one of the largest public bike hire systems in the world. The London Cycle Hire Scheme uses almost 12,000 bikes and 800 docking stations and accounts for more than 10 million rides per year. Decision Brain uses Machine Learning tools for the real-time forecasting of user demands, and applies optimization algorithms for inventory management and bike relocation. Their algorithms make sure that when you arrive at a station, there are sufficient bikes available to rent and sufficient empty slots for bike return.

When do we need to keep humans in the loop?

As demonstrated by the above examples, there is no doubt that AI-powered analytics has enormous potential for leveraging the competitive advantage of businesses or helping them run sustainable and environmentally-friendly operations.

However, there are some important decisions where we do not want to let algorithms fully replace human judgment. An obvious example is the case of long-term decisions that may involve many stakeholders or require a careful analysis of geopolitical or other strategic aspects. In this case,


© GettyImages - Metamorworks

the data-driven AI+BA approach can provide analysis of possible outcomes for several future scenarios (worst-case, expected, and best-case, for example). However, these outcomes can be only seen as recommendations for the executive board, not as decisions to be implemented immediately.

Similarly, automatic decision making may lead to unintended consequences, especially when it comes to sensitive policy making that profoundly affects human lives. It is very tempting to let AI analyze the large amount of available data to produce various metrics and turn people's lives into scores that predict ethically ambiguous outcomes like the criminal potential of certain neighborhoods (predictive policing), employee productivity (AI for HR), or recidivism (when making sentencing decisions at the court). However, these algorithms are highly sensitive to the input data and as such, they can easily amplify discrimination and be turned into "Weapons of Math Destruction", as explained by Cathy O'Neil in her book of the same title.

However, it is important to know that many algorithms are made with best intentions and no algorithm is evil in itself! The algorithms may encode human prejudice or misunderstanding, and in our data-empowered and accelerated economies, this can lead towards increased inequalities or perpetuation of the racial and gender biases with unpredictable consequences.

This is why business leaders today need to understand challenges, opportunities and limitations of AI and analytics. New generations of managers at ESSEC are learning coding - not to become software engineers or data scientists, but to understand the "algorithmic thinking" and the importance of developing explainable and ethical decision-making tools based on AI.

WHO SHOULD OWN YOUR DATA? AN ECONOMIC ANALYSIS

Wilfried Sand-Zantman *joined ESSEC Business School in 2020 and was previously Professor of Economics at Toulouse School of Economics. He holds a Master in Economics and Statistics from the ENSAE (Paris) and a PhD in Economics from the University of Toulouse 1. Wilfried's research interests are in industrial organization, digital economics, and more generally informational economics. His articles have been published in journals such as the International Journal of Industrial Organization, the Journal of Industrial Economics, Management Science, the Rand Journal of Economics, and the Journal of European Economic Association. Wilfried was also involved in consulting and executive education for several firms as EDF, Orange, SFR and Telefonica.*

Data has become a major economic issue. In the digital ecosystem, the harvesting, processing and resale of information compose the strategy of many firms. Whether by providing a free service in exchange for attention or by collecting data during a transaction, many firms use the data provided by consumers, the latter being often not fully aware of it. To the extent that this data is an asset in the production process, the question of control over this data and thus ownership of the data arises.

In Europe (with the General Data Protection Regulation) as in the United States (with the California Consumer Privacy Act), there are legal responses that allow consumers to better control the collection and use of their data. But the question of ownership, and therefore the possibility of directly monetizing personal data, is still approached in a roundabout way (See Duch-Brown et al., 2017). Paradoxically, we've also seen a rise of data brokers who operate in an active data market (see FTC, 2014) and nothing prevents an individual from selling his/her private data in exchange for remuneration. However, data cannot really be transferred in the traditional way because the consumer can prevent a third party from using it (at least in Europe), even if he or she has previously authorized its use.

Despite these practical difficulties, both legal scholars and economists are increasingly tackling the topic. Beyond studying the consequences of implementing a data market, we need to understand the impact of allocating property rights either to consumers or to the firms that have made it possible to extract the data. The starting point is the idea that data as an economic object is the result of an interaction between two parties. Data has a special status as both input and output. Like other goods, allocating the rights to one party or the other changes the way people consume or produce (see Coase 1960).

In our recent paper (Dosis and Sand-Zantman, 2019), we propose a theoretical exploration of this issue in a two-sided market framework. More precisely, we analyze a situation in which consumers consume a service, with the data generated during this transaction able to be monetized subsequently (personalized offers, sale to data brokers, internal use, etc.). Two important assumptions are made in this study. Firstly, the market value of a given

© GettyImages - NicoElNino

customer's data is all the greater the more he or she uses the service and the more these data have been processed by the firm. Second, consumers are concerned about their data being exploited, which negatively impacts their satisfaction when using the service. In this framework, we suggest a trade-off between two forms of inefficiency, with overexploitation of data on one hand and data under-processing on the other.

When firms own the rights to their consumers' data, as is currently the de facto case in most countries, they have the opportunity to monetize it and therefore have an incentive to both generate a lot of it and to process it efficiently to generate the greatest value. However, consumers are aware of this risk and tend to restrict their use of the service in question, particularly when they would have liked to use it intensively absent this risk. This results in an efficient data processing but limited use of the service by consumers.

When consumers have the control rights over their data instead, and can trade it on a market, they can adjust their decision to trade or not by considering the full benefits and costs.. On the flipside, this means that firms have no reason to process the data efficiently, which limits the money they can obtain on the market for data when they trade.

So what is the optimal choice between these two ownership regimes? Even if the interests of firms and consumers are not totally aligned, it can be shown that they evolve in the same direction when the data value on the second market - the one in which the data is resold - changes.

Specifically, if the market value of the data in the second market is low, firms gain little from processing it fully. But the exploitation of data, or the risk of exploitation, reduces the intensity of use of the consumption of the service, and thus the firm's direct income on this market. On the contrary, leaving the ownership of the data to the consumers guarantees a reasonable use of the data, at a limited opportunity cost since their market value is low. If, on the other hand, the market value of data is high, it is important for firms to exploit it monetarily, and to process it to extract the maximum value from it. This also leads them to propose very advantageous offers to consumers, if they have the possibility of adding value to the data generated.

In other words, when data has little market value, the inefficiencies associated with under-investment by firms in the valuation process are small relative to the benefits to consumers of having more control over the use of their data. On the other hand, where data has significant market value, granting firms the right to exploit consumer data leads to the creation of enough value to compensate consumers for the inconvenience of using their data.

Beyond questions about the relevance of monetizing personal data, this approach leaves important questions unanswered. First, determining the value of personal data is complex. Data is used internally (to improve the service offered) but also externally (for targeting or to be sold on a market) and determining the value of data for a firm is a very speculative task. Second, the value of an agent's personal data does not only depend on that agent. Indeed, information about one person can sometimes be retrieved from the personal data of others, such as friends, consumers with the same profile, etc. (see Choi et al., 2019). Finally, the economic transactions generating an individual's data may involve several firms at the same time or the same data may be generated in parallel by independent economic transactions. The question of multiple ownership, or exclusive ownership, then arises, making the constitution of a market for data even more complex. ■

References

Choi, J.P., Jeon, D.-S., & Kim, B.-C., (2019). Privacy and data collection with information externalities. *Journal of Public Economics,* 173, 113-124.

Coase, R. (1960). The problem of Social Cost. *Journal of Law and Economics,* 4, 1-44.

Dosis, A., & Sand-Zantman, W. (2019). The Ownership of Data. mimeo TSE and ESSEC.

Duch-Brown, N., Martens, B., & Mueller-Langer, F. (2017). The Economics of Ownership, Access and Trade in Digital Data. JRC Digital Economy, Working Paper 2017- 01.

Federal Trade Commission (2014). Data brokers: A call for transparency and accountability. Report. TC, Washington DC.

© GettyImages - Sdecoret

DO WE STILL NEED RESEARCHERS?

__Vivianna Fang He__ is Associate Professor of Management. Prof. He studies collaboration in three innovation-related contexts: new venture teams, research and development projects, and open-source software communities. Her research, which combines organizational and psychological theories in novel ways, has been published in leading international journals and won competitive national grants.

In the age of AI, do we still need researchers?

Do we still need researchers given the rise and reach of artificial intelligence? We know that machine learning can identify complex relationships in massive datasets that are not necessarily identifiable to the "naked" (human) eye. We also know that artificial intelligence will eventually be able to take over many human functions. Will research be one of them? Vivianna Fang He of ESSEC Business School and her colleagues Yash Raj Shreshta (ETH Zurich), Phanish Puranam (INSEAD), and Georg von Krogh (ETH Zurich) dive into this question in their recent research.

In a word: no, but there's more to it than that.

Until now, machine learning techniques have been widely used for coding data and making predictions, but not yet for the core task of a researcher--building theory. Why is this? It could be due to a scholarly distaste for so-called "predictions without explanations". In fact, this is exactly where the opportunity lies, Prof. He and colleagues suggest. Machine learning could indeed do a better job than researchers in finding robust and complex patterns in the data.

Traditionally, organizational scientists propose a theory and/or model and then test it, usually using a relatively small dataset. With larger datasets, there's more of a chance that the results will be applicable to a wider population rather than just the one used in the study and thus be *replicable*--true in other situations other than the one at hand. Researchers can also study more variables when working with larger datasets, which is invaluable for constructing a more complete picture of the situation we are studying. When building a theory from data using traditional statistical tools, researchers run the risk of *overfitting*--finding a pattern that is specific to the current sample. Machine learning algorithms have procedures that help avoid overfitting, meaning that the patterns they identify are more likely to be reproduced in other samples. This is a highly valuable property, as it could, for example, help address psychology's current replication crisis by facilitating the development of robust theories and therefore replicable results.

© GettyImages - Igor Kutyaev

Another advantage of integrating machine learning techniques is that it can help manage researcher bias by making the research processes and decisions transparent. Researchers are only human, after all, so it is possible that they'll experience confirmation bias and look for results that support their predictions: in other words, that they'll see what they want to see. In using machine learning algorithms, researchers can specify the level of complexity in the patterns detected and document these decisions. Together these procedures allow for a thoughtful approach to balancing predictive accuracy and interpretability. Higher predictive accuracy can mean that the patterns are too complex to understand, and higher interpretability can mean that the pattern is simpler and perhaps then not taking all impactful factors into account. Being able to control this tradeoff is essential for interpreting the patterns in a way that makes sense to people and not just to machines. It also means that researchers can explain their rationale in a transparent way.

However, the machines can't act alone: algorithms lack the intuition and common sense that humans have. While they can put the pieces of the puzzle together, it is up to us humans to explain why the pieces go together. Many critical parts of the theory building process will still be up to the researchers, such as defining what factors are of interest, selecting or developing ways to measure those factors, and explaining the relationships driving the observed patterns. The future of theorizing will require a synergy between algorithms and humans.

Prof. He and her colleagues propose a four-stage procedure to explore this opportunity. The first stage is splitting the sample into two: one sample to use for machine learning-supported pattern detection, and the second sample to use for testing the hypotheses. In stage 2, the researchers program the algorithms and the algorithms do their magic and identify interpretable and reliable patterns. In stage 3, the researchers ask themselves if the patterns make sense, and come up with the explanations for the patterns. This stage is where human expertise and judgement are essential, as ML algorithms don't have the capacity to do this. In stage 4, researchers test the hypotheses- the theory- in the second sample to see if the pattern holds.

The authors have applied this method to study the governance disputes in online communities (He, Puranam, Shreshta, von Krogh, 2020), whereas other organizational scholars have used it to identify optimal revenue for a wide range of App store products (Tidhar & Eisenhardt, 2020) and to gauge whether or not an idea will take off (Dahlander, Fenger, Beretta, Kazami, & Frederiksen, 2020). Similar approaches are also being experimented in natural sciences. For example, Udrescu and Tegmark (2020), two physicists at MIT, used 100 equations to generate data and then feed that data to a neural network. Their algorithm was able to recover all 100 equations! This diverse set of studies show that the approach can be applied to a wide variety of topics, making it useful for researchers across disciplines.

While this approach has extensive implications for theory-building, the authors do note that there are some caveats to be considered before using this approach. Machine learning assumes that the future can be predicted from the past, so it's best to use machine learning algorithms when assessing relatively stable phenomena. Second, machine learning cannot replace randomization. Machine learning techniques are most suitable for coming up with predictions, rather than testing a theory about the relationships between variables.

There is also the risk that ML techniques could amplify biases present in the data, leading to biased conclusions, as biases could be hard to detect but have significant ethical consequences. Therefore, researchers must have a strong conceptual understanding of the techniques they're using, which is no easy feat in such a rapidly advancing field.

In a nutshell, while machine learning cannot replace researchers, it CAN take over some of the functions that humans currently do, like pattern recognition, rote memorization, and arithmetic. However, people are needed for tasks that require more intuition and creativity, like explaining patterns, book writing, and art.

So, do we still need researchers? Yes - and machine learning can be a powerful tool for producing more robust research.

For more information on using machine learning algorithms in theory-building, check out their article here: https://pubsonline.informs.org/doi/full/10.1287/orsc.2020.1382

References

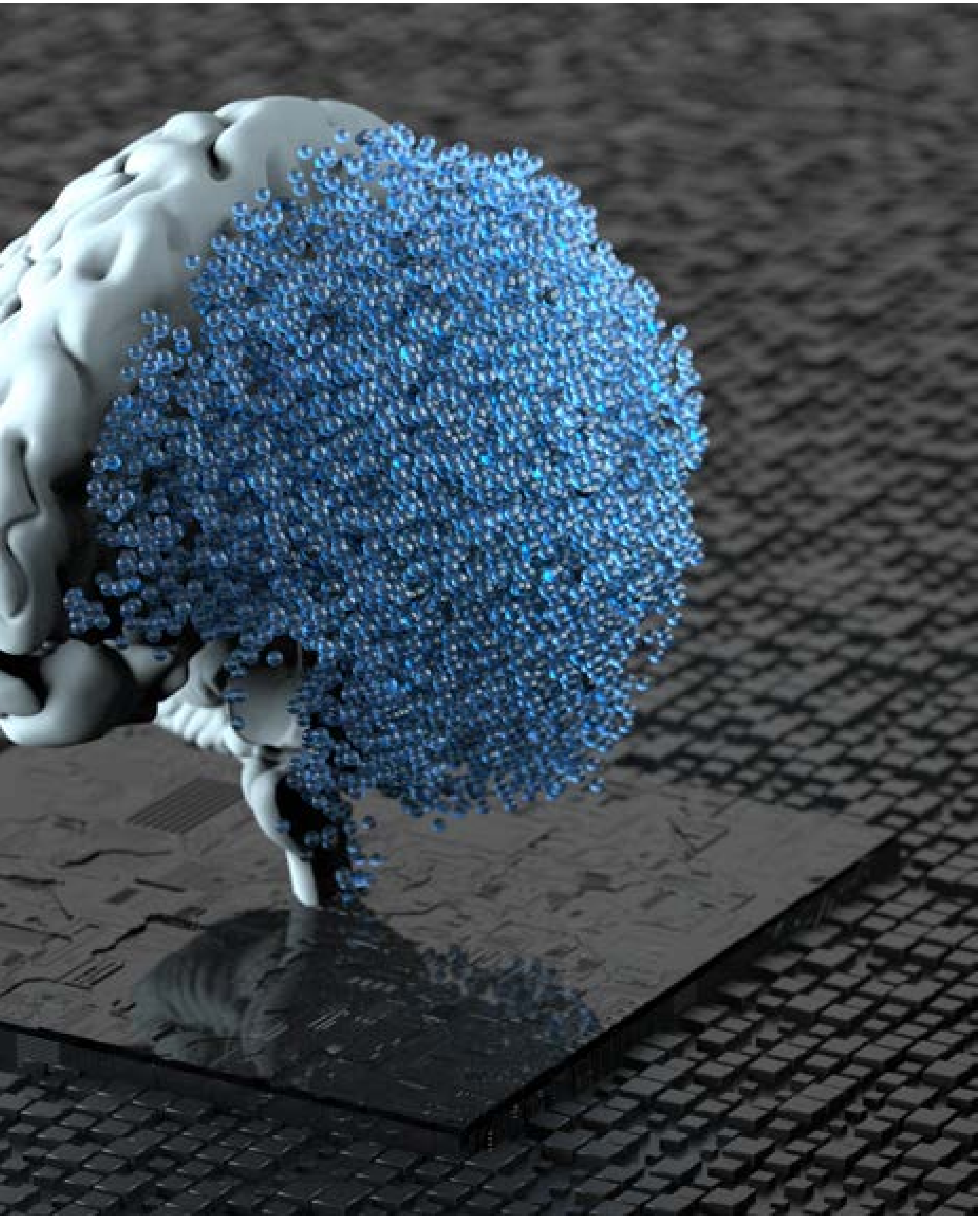
Dahlander, L., Fenger, M., Beretta, M., Kazami, S., & Frederiksen, L. (2020). A Machine-Learning Approach to Creative Forecasting of Ideas. In *Academy of Management Proceedings* (Vol. 2020, No. 1, p. 17027). Briarcliff Manor, NY 10510: Academy of Management.

He, V. F., Puranam, P., Shrestha, Y. R., & von Krogh, G. (2020). Resolving governance disputes in communities: A study of software license decisions. *Strategic Management Journal,* 41, doi: 10.1002/smj.3181

Shrestha, Y. R., He, V. F., Puranam, P. and von Krogh, G., (Forthcoming) Algorithm supported induction for building theory: How can we use prediction models to theorize? *Organization Science.*

Tidhar, R, Eisenhardt, KM. (2020). Get rich or die trying… finding revenue model fit using machine learning and multiple cases. *Strategic Management Journal,* 41: 1245– 1273. Doi : 10.1002/smj.3142

Udrescu, S. M., & Tegmark, M. (2020). AI Feynman: A physics-inspired method for symbolic regression. *Science Advances,* 6(16), eaay2631. DOI: 10.1126/sciadv.aay2631



© GettyImages - Just Super

ARTIFICIAL INTELLIGENCE AND INFORMATION GOVERNANCE: A PROMISING PAIR?

Isabelle Comyn-Wattiau *is Professor at ESSEC Business School, where she teaches information systems. She founded the chair Information Strategy and Governance. Her current research projects explore information strategy and governance, digital humanities, modeling and design of information systems. Her research has been published in many leading academic journals including the Journal of Management Information Systems, Decision Support Systems, Expert Systems with Applications, Data and Knowledge Engineering, and few others. She chairs the Steering Committee of the International Conference on Conceptual Modeling. She serves on the Editorial Review Board of several academic journals, including Data and Knowledge Engineering and Systèmes d'information et management. She was Editor-in-Chief of Ingénierie des Systèmes d'Information (Lavoisier) and recently created the open journal Revue Ouverte d'Ingénierie des Systèmes d'Information.*

Artificial intelligence has taken the business world by storm, bringing changes in decision-making supports. It is a new addition to the information component toolkit that all organizations must implement (or at least consider). The introduction of the General Data Protection Regulation (GDPR) in Europe in 2018 and its equivalents in other countries and continents have highlighted the challenge that proper information governance poses to organizations.. In this article, we tackle two elements of the new situation: artificial intelligence and information governance. After a refresher on what information governance entails, we will address two questions: how must information governance take into account the new artificial intelligence tools, which are both consumers and producers of information that needs to be governed? How can these same tools help organizations with data and information governance?

1. What is information governance?

Information governance refers to the establishment, in a company or organization, of an information management policy, with defined objectives and a planned implementation, including human, organizational, and technological resources. The terms "data governance" and "information governance" are used more or less interchangeably to refer to this kind of information asset management strategy. Using the term "data governance" emphasizes the digital aspect of the strategy, but may limit the scope to structured data. Information governance encapsulates all forms of data, regardless of their form: structured data in tables, semi-structured data in documents, or less structured data in messaging software or images. Ultimately, we can define information governance by its objective: maximizing the value of information while minimizing the associated costs and risks.

Information governance includes a set of key processes, such as managing data quality and security (availability, integrity, confidentiality, traceability). While these

© GettyImages - Artem Peretiatko

processes often use IT processes, they are often the responsibility of other business units, and rightfully so.

Information governance is relevant for all areas and sectors of activity. It is especially pertinent when the three complementary dimensions of value, cost, and risk are significant. For example, take the banking sector, where the cost of information management can represent over 10% of revenues and the associated risk is significant, as it can involve, for example, handling sensitive information like money in electronic form! In the health sector, the risk – meaning information disclosure or error – is the most important feature, though we cannot ignore the other dimensions.

Information governance is increasingly better organized in public and private companies, and more broadly, in all organizations. The latter are becoming more advanced, as indicated in other studies. That said, they must constantly face new challenges, including artificial intelligence.

2. The singular experience of governing artificial intelligence

Artificial intelligence tools store data and information to understand, decide, and learn. Data and information are increasingly voluminous and varied. As with all the company's information, they must be under control, which is one of the key aspects of information governance. Be it data warehouses, data lakes, or another way to store information, nothing must be left off of the information map and all must fit into the dimensions of valuation, risk management, and cost control.

Data quality must live up to its name: it must facilitate the use of artificial intelligence by rendering the processes of collecting, locating, and cleaning data readily available.

Above and beyond these classic aspects of information governance, artificial intelligence raises new questions, like the presence of human biases in both algorithms and data and how this could lead to poor decisions. In addition to collecting personal data, and more generally, sensitive data, the use of information gathered by artificial intelligence must also include an explanation of these algorithms.

This is the challenge of explainable AI, which refers to the methods that equip the "black boxes" of AI with modules that at least partially explain the results they provide. In many cases, this becomes a regulatory challenge. It can also be a trust issue for users to rely on the precious aid AI can provide without reluctance. Finally, it is an ethical challenge for organizations, who are thus able to ensure that their processes are transparent, even when they use artificial intelligence tools.

Artificial intelligence at the service of information governance

Conversely, artificial intelligence tools could also reveal themselves to be a precious tool for information governance.

Therefore, mapping information must, among other things, identify personal data to ensure compliance with the GDPR. Artificial intelligence could help maintain this map, a tedious but important task, which requires analysis and decision-making skills. The volume and omnipresence of data makes this task impossible to accomplish manually.

Humans will remain the masters of the information governance ecosystem, but artificial intelligence and other emerging technologies, such as blockchain, will play a very important role in helping analyze information and ensure compliance. Automation could also help with archiving and destroying information, the ultimate phase of the information life cycle, thus addressing both regulatory and efficiency needs.

Software editors who specialize in information governance have reported integrating artificial intelligence into their component offering, though it remains difficult to evaluate the scope of this integration. Some use intelligence techniques to identify new data, uncover relationships in the data, and classify documents. Learning techniques make it possible to describe the type of content that the software can recognize after sufficient training. In natural language processing, a technique of artificial intelligence, the automatic recognition of characters is increasingly improving. Software robots can execute repetitive tasks, like locating bills in a set of administrative documents and providing the human with an analysis that he or she only needs to confirm, like how ATMs can process cheques automatically.

Explainable AI would also help the organization demonstrate its strong information management when dealing with regulators or auditors. Finally, organizations must face regulations that are constantly changing and that vary from one country to another, in a world with increasingly rich and varied data: in this context, information governance has much to gain from making use of artificial intelligence tools. ■

© GettyImages - Metamorworks

AI FOR PERSONAL DEFAULT PREDICTION: PROMISES AND CHALLENGES AHEAD

Andras Fulop *is Professor of Finance. He teaches Empirical Asset Pricing and Financial Risk Management in the PhD, Master in Finance and Master in Management programs. Andras's research interests are broadly in asset pricing with an emphasis on credit risk, financial econometrics and derivatives. He has published in academic journals such as* Review of Financial Studies, Journal of Econometrics *and* Journal of Business and Economic Statistics. *He holds a Ph.D., in Finance from Rotman School of Management, University of Toronto.*

Laurence Daures-Lescourret *is Associate Professor of Finance. Her research lies in the area of microstructure of financial markets. Laurence received the PhD Thesis Award in 2004 from the French Finance Association and the National Foundation for Companies Management Academic Education (FNEGE), the "Joseph de la Vega Prize" in 2013 and the IFSID award for the best paper on Derivatives in 2015. She received several research grants from Euronext Paris (2007), EIF (2008, 2010, 2020) , the French National Research Agency (JCJC, 2011) and INEX (Initiative d'Excellence, 2018). She serves on the Board of Directors of Dassault Systèmes and of LCL (Le Crédit Lyonnais). She sits in the audit committee of Dassault Systèmes (DS) and chairs the compensation and nomination committee of DS. She sits in the risk committee of LCL and chairs the audit committee of LCL. Laurence holds a PhD in Finance from HEC Paris.*

Information technology has been shaping consumer credit risk for decades. The Fair Isaac Corporation (FICO) introduced FICO scores in 1989, marking a milestone in moving the credit evaluation process from humans towards algorithms. The FICO score is a credit score that takes into account five areas to determine creditworthiness: payment history, current level of indebtedness, types of credit used, length of credit history, and new credit accounts. Today, it is used in more than 90% of the credit decisions made in the U.S.

With the increasing digitalization of society in the last decade, there has been an explosion both in the collection of personal data and in the sophistication of the algorithms and computing capacity to process all this information. This clearly holds the potential to fundamentally impact the process of evaluating individuals' creditworthiness. In this piece, we summarize some of the lessons that can be gleaned from the recent academic literature arising from the application of powerful artificial intelligence (AI) techniques to consumer credit risk.

© GettyImages - Tolgart

How can AI help?

The general finding is that newer AI tools indeed live up to their promise of improving the technology of credit screening. Traditional credit scoring models such as the one behind the FICO score are based on hard information sources from individuals' financial accounts. A relatively small number of predictive variables are usually then combined using rating scorecards, or linear models such as logit regressions. Newer AI approaches go beyond such tools in at least two important aspects.

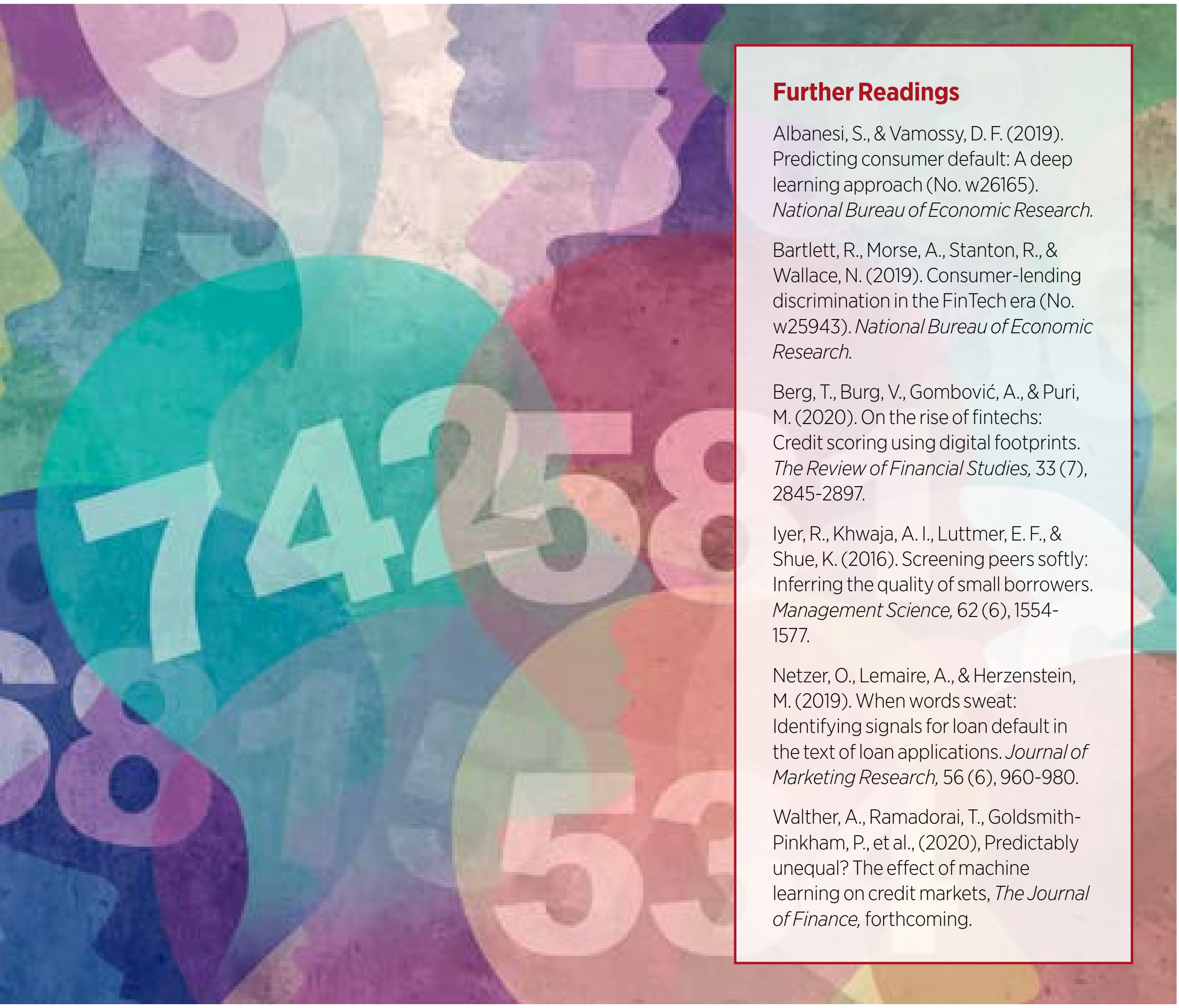
First, modern machine learning algorithms, such as tree-based methods or neural nets, allow flexible non-linear predictive relationships between predictors and the individuals' credit risk, while tackling in-sample overfitting through sophisticated regularization methods. The overall message from the literature is that machine learning techniques tend to outperform traditional linear models, especially within higher risk groups. For instance, Walther et al. (2020), using a dataset of millions of US mortgages, showed that tree-based methods significantly outperform logit techniques. Albanesi and Vamossy (2019) compared several machine learning algorithms on data from the Experian credit bureau and found that their ensemble technique combining neural networks and gradient-boosted trees improves upon traditional models and the improvement is especially pronounced for consumers with low credit scores.

Second, digitalization and AI algorithms allow the use of new types of data. Using a dataset covering around 270,000 purchases at a German e-commerce company, Berg et al. (2020) analysed the predictive power of digital footprints (the information left behind by individuals while visiting a website, such as the device, the email address, the shopping hour, etc.). They found that the accuracy of a credit risk model built on digital footprint variables is comparable and complementary to that of traditional credit risk scores. An important implication of this result is that digital footprints can help in screening some borrowers with little credit history. Another source of information that AI algorithms can benefit from is unstructured user data in the form of text or images (for instance, information from social networks such as LinkedIn, Twitter or Facebook). Using data from Prosper, a crowdfunding platform, Netzer et al. (2019) find that supplementing the financial and demographic information with the textual information submitted by prospective borrowers substantially improves default prediction. Using a similar dataset, Iyer et al. (2016) found that the market interest rate that mirrors the information available to lenders on Prosper is a more precise predictor of defaults than traditional credit scores.

What are the implications of these new techniques on consumer welfare?

On a positive note, improvements in screening technology that are particularly pronounced among riskier groups and people with scant credit history [Berg et al (2020)] can decrease asymmetric information problems between borrowers and lenders and lead to increased access to credit. This feature can be especially useful in emerging countries with limited reach of the formal banking sector and hence a lack of traditional credit information for most consumers.

A further potential benefit of turning credit decisions over to algorithms is that human biases of loan officers – such as racism and in-group/out-group bias – can be short-circuited, leading to less discrimination. To examine this issue empirically, one first needs a precise definition of discrimination. Bartlett et al (2019) suggested using the interpretation of US courts, whereby any differential impact of the treatment of minority groups not related to 'legitimate-business-necessity' is deemed discriminatory. In the credit context, 'legitimate-business-necessity' essentially means variables that help in predicting default risk. Hence, to measure discrimination against minority groups, one would need to compare consumers from minority groups with peers from majority groups and the same credit risk. Given that empirically the credit risk of individuals is observed only imperfectly, this leads to an omitted variable problem. Bartlett et al. (2019) used an identification afforded by the pricing of mortgage credit risk by government-sponsored entities (GSEs) (Fannie Mae and Freddie Mac) in the US to deal with this issue. In particular, these GSEs use a predetermined grid that prices credit risk across loan-to-value and credit-score buckets. Given that the credit risk of conforming mortgages are insured by these GSEs, any access or price differences for borrowers within the same bucket are unrelated to creditworthiness. and fail to qualify as 'legitimate business necessities' but qualify as discrimination. Using this empirical strategy, Bartlett et al. (2019) found that FinTech algorithms also discriminate, but 40% less than face-to-face lenders in pricing mortgages. What's more, FinTechs do not discriminate in loan approval, while face-to-face lenders do.



© GettyImages - Wildpixel

Further Readings

Albanesi, S., & Vamossy, D. F. (2019). Predicting consumer default: A deep learning approach (No. w26165). *National Bureau of Economic Research.*

Bartlett, R., Morse, A., Stanton, R., & Wallace, N. (2019). Consumer-lending discrimination in the FinTech era (No. w25943). *National Bureau of Economic Research.*

Berg, T., Burg, V., Gombović, A., & Puri, M. (2020). On the rise of fintechs: Credit scoring using digital footprints. *The Review of Financial Studies,* 33 (7), 2845-2897.

Iyer, R., Khwaja, A. I., Luttmer, E. F., & Shue, K. (2016). Screening peers softly: Inferring the quality of small borrowers. *Management Science,* 62 (6), 1554-1577.

Netzer, O., Lemaire, A., & Herzenstein, M. (2019). When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research,* 56 (6), 960-980.

Walther, A., Ramadorai, T., Goldsmith-Pinkham, P., et al., (2020), Predictably unequal? The effect of machine learning on credit markets, *The Journal of Finance,* forthcoming.

New challenges of AI algorithms

First, more precise screening algorithms tend to lead to more inequality in the cost of credit among consumers. This increase in dispersion is particularly pronounced among minorities and riskier borrowers [see Walther et al (2020)]. Shaping policies that improve the terms of credit for disadvantaged households in the presence of such improved screening technology is a crucial topic for the regulatory debate going forward.

Second, as Bartlett et al. (2019) pointed out, AI algorithms may also increase the performance of screening of consumers in non-'legitimate-business-necessity' dimensions. In particular, if a lender uses such algorithms to maximize profits unrelated to screening for credit risk and such profit-maximizing screening has a differential impact on protected minority groups, the company risks coming under the purview of anti-discriminatory legislation even if there is no personal bias against minorities in the algorithm. Further, the black-box nature of most AI algorithms increases the risk of such scenarios, as its functioning may not be clear to the humans operating it. Hence, a key challenge is the development of non-discriminatory AI algorithms. The future of AI algorithms in credit decisions is bright, but its human operators must be careful to ensure that they understand how it works and aim to reduce the risk of inequitable decisions to provide fair, accurate evaluations to all. ■

THE LINK BETWEEN DATA MATURITY, DATA TALENTS, AND RENEWABLE ENERGY

Jeroen Rombouts *has been a professor at ESSEC Business School since 2013. He holds the Accenture Strategic Business Analytics chair. He obtained his PhD in econometrics in 2004. As a full professor of statistics and econometrics, he teaches big data analytics and data science courses at the MBA and Master's programs in Paris and Singapore. He researches predictive analytics, in particular time series analysis and forecasting. He has published numerous scientific articles, frequently organizes workshops, and is associate editor of several quantitative scientific journals. His research is funded by the European Commission and the ANR. He also consults on strategic business analytics and data science topics.*

Co-author information
Gérard Guinamand: *Gérard Guinamand is the Group Chief Data Officer at Engie.*

Over the last few years, the web giants have shown that using data to know your customers is key for the development of new products and services, and for beating the competition. These companies operate on a digital core, allowing data-augmented and data-driven decision-making, and are highly appreciated by investors given their massive market capitalisations. Even during the COVID-19 pandemic, the tech sector continued growing spectacularly given the acceleration in digitalising the way we work and interact with others.

Any large, mature company is inspired by the way tech companies operate and dominate. For example, in France, L'Oréal aims to become the top beauty tech company by using artificial intelligence and augmented reality. Since 2018, Carrefour has used the Carrefour-Google Lab to accelerate its digital transformation. Danone and Microsoft launched The AI Factory for Agrifood in 2020. Energy companies like Engie and EDF are pushed by the general public sentiment on climate change to become operationally excellent and greener. Young data talents are hired to help transform the companies and introduce the new data culture.

Predictive mindset

In theory, the smart use of data and the creation of business value makes a lot of sense, though in practice traditional companies struggle within becoming more data-driven. Companies have been investing largely in data infrastructure over the last years, appointed chief data officers, and launched data training programs to convince every employee of the salient features of data and analytics. Consequently, massive amounts of data are stored in the cloud and often the question now is "what can we do with this?" or "what is the actual return on all these data investments?". To answer such questions, a next step in the data maturity process of the company is essential. This is the step towards becoming more data- informed, data-driven, and operationally excellent, and it requires using data to look forward, rather than backward, and therefore make predictions. To put it differently, after storing and categorising data, it is now time to use it for decision-making across all levels, rather than specific pockets, of the organisation. In fact, business decisions always implicitly include predictions, and it is time to make this process more formal and automatic thanks to the use of data.

© GettyImages - Blue Planet Studio

The predictive paradigm is not only about recommendation algorithms and the like: it also allows for the use of data at the highest executive level to ensure that strategy is implemented. Specifically, the management of forward-looking key performance indicators (KPIs) allows for measuring and tracking the success of the company and setting clear objectives. This in itself generates valuable data that can be correlated with new initiatives to predict their success and gain deeper understanding of their link with existing operations. To sum it up, C-level executives need to start implementing strategy with data rather than strategy for data, so that the company's operating model can become data-centric in the same way as famous tech companies like Amazon and Ali Baba.

Someone once said, "Making predictions is hard, especially about the future". Predictions are by nature uncertain and this has to be incorporated when making business decisions, similar to financial investors that use more information than the average return on an asset for deciding to buy or sell it. Accurate predictions are obtained when combining varying sorts of data, including external sources like weather-related data in energy applications. How to actually produce predictions using data is not a trivial task. It demands talented data scientists and advanced algorithms, plus continuous performance monitoring. It is not a surprise that the International Data Corporation expects global spending on artificial intelligence to increase from 43 billion EUR to 94 billion by 2024.

Renewables

The International Energy Agency expects renewables to provide 80% of the growth in global electricity demand through 2030. In fact, solar- and wind-energy projects have become less expensive, and interest rates are historically low today. Furthermore, governments are highly supportive, exemplified by the European Green Deal, whose main ambition is making the EU climate neutral by 2050. Renewable energies are therefore becoming a key strategic goal for the energy companies. Technically speaking, renewable energies (in particular wind and solar) have a high level of intermittence (night, absence of wind) but we cannot increase the number of plants to compensate for this lack of production for economic and environmental reasons. This situation implies two main actions for energy companies such as ENGIE. First, identify the best sites for new implementations. Second, get the best performance from the plants while taking into account operating constraints (noise for example) for a higher volume of electricity generated.

Data plays an essential role for the success of renewable development because iIt enables: the best selection of the optimal sites based on topography and weather forecast data,; the best results based on technical availability, real-time weather and measurement data related to environmental constraints (noise), and optimizing of electricity sales by combining production data with data on demand, market prices, and storage capacities.

In terms of return on data investments, renewable energy forms a perfect use case where the value from data can be made explicit. In fact, these sources of energy are highly sensor equipped allowing for predictive maintenance, and for data monetisation. On the production side, there are no GDPR concerns. Operational excellence in the renewable energy business will be the only way to survive for incumbents. Traditional oil companies such as B.P. and Total are rapidly transforming themselves and will compete fiercely with the current energy players in the market of renewable energy.

ENGIE - ESSEC

Engie and ESSEC Business School have been working on different cases for three years as part of the Strategic Business Analytics Chair sponsored by Accenture. The Chair's main objective is to train the next generation of leaders to develop new business strategies, leveraging the numerous applications of advanced analytics. Through a hybrid learning method based on innovation, collaboration and entrepreneurship, the Chair acts as the core of an ecosystem combining data and value creation – from purpose and strategy crafting to transformation, encompassing problem solving, data science & artificial intelligence, culture change and skills development.

Engie is an important part of the Strategic Business Analytics Chair's ecosystem. In 2021, the Chair students will work on two strategic cases on renewable energy. Their fresh and forward-looking vision on the topics generates innovative ideas and valuable solutions. ESSEC students are particularly interested in working with companies like Engie given its strong environmentally-oriented strategic values. Indeed, the students, being concerned about climate change, prefer to work on business cases that ultimately generate societal value rather than purely commercial cases for e-commerce platforms.

Looking forward

In the future, renewable energy will lead to the creation of many new jobs requiring technical, data and analytics skills. The EU reports that already the solar photovoltaic industry alone accounted for 81,000 jobs with expected increase to 175,000 and 200,000-300,000 jobs in 2021 and 2030 respectively. Digitalisation and renewable energy go hand in hand and will be an important driver for economic growth. The Partnership with Engie and Essec ESSEC will guarantee that young talents are trained and acquire the skills to make sure the transition to a green society is completed satisfying the climate change agreements.

More generally, it will be the companies which employ the people with the right skills, mindset and vision that will make the difference. Data is now available, most analytics tools that create value are standard, and computational resources have little constraints. It is the culture of the company that requires a fundamental change. Those who will be able to attract young "data ready" business graduates are going to be at the competitive edge.

Accenture mentions in its Technology Vision 2020 that the tech-clash is a new situation, where on one hand, people are enthusiastic about technology, data and artificial intelligence, but on the other hand, they require algorithms to be understandable and fair, and know where their personal data is used. This balance will be extremely important in the post COVID-19 era, where all that matters will be human experiences.

© GettyImages - Blue Planet Studio

THE ETHICS OF ARTIFICIAL INTELLIGENCE

Laurent Bibard *is Professor of Management at ESSEC Business School. He holds PhDs in both Socio-economics (EHESS) and Philosophy (Université Paris IV Sorbonne). Laurent was at the head of the Edgar Morin Chair on Complexity from 2014-2019. His research examines gender, focusing on gender-related political stakes, as well as on how high reliability organizations manage for resilience.*

When pondering the ethical questions posed by artificial intelligence, it's important to keep two main points in mind:

1. Developing new technologies and advancing artificial intelligence is not an ethical issue in and of itself. It is not technology that poses the problem, but rather how we use it and how we dream of using it. This clarification is a tale as old as time, and it is crucial to remember it and to combat our naïveté when debating the ethics of artificial intelligence.

2. Today, this question appears in countless different forms and scenarios. This showcases the fascination inspired by "new technologies" and shows that artificial intelligence has progressed immensely. The biggest problem is actually that it is implicitly assumed or explicitly stated that artificial intelligence will not only surpass human intelligence in the future, and even that it has already done so. Among system designers, this results in the assumption that humans are ultimately only a source of error. One of the major consequences of this assumption are that the engineers who design the systems that are supposed to replace humans are *de facto,* without question, far superior in all senses – including ethically – to the humans who are supposed to manage said systems. This is particularly true when one is supposed to anticipate the ethical decision-making problems in potential emergency situations, such as for self-driving vehicles or in aviation.

Putting aside the grave ethical problems linked to short-term thinking and cost-cutting that Boeing, for example, experienced in the case of its rushed conception of the Boeing 737 Max, there is also an ethical problem specifically linked to its design, implementation and the (dys)functions of the MCAS system meant to ensure aircraft safety in the event of stalling due to a loss of speed.

When an airplane is losing speed and at risk of stalling, the MCAS system is supposed to induce its path downward to make it regain speed and recover altitude. An airplane risks losing speed when it pitches up, which is what tragically happened to the plane in a 2009 Air France Rio-Paris flight. The MCAS software was designed to automatically make the plane dive

downward when the data signaled to the system that the plane was tilting too high and at risk of stalling. This is done without involvement of the pilots, as their vigilance and effectiveness was considered inferior to that of the automatic system.

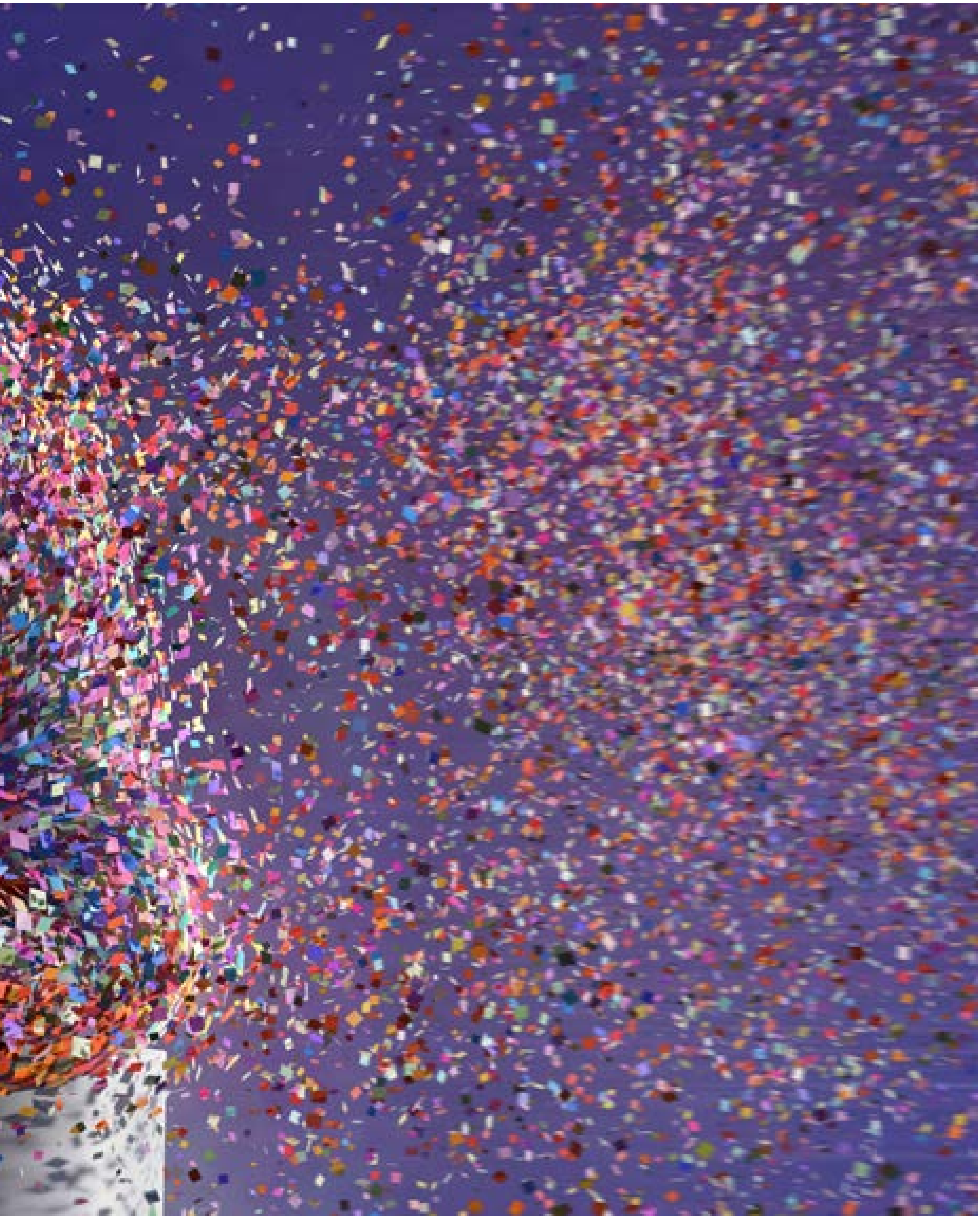
But here's the thing. If the system misinterprets the data, the system could then "interpret" that the plane is tilting up unnecessarily, even though it might be in a normal – and therefore essential – ascension phase after takeoff. This is what happened in the two catastrophic cases of the Southwest Airlines and Ethiopian Airlines in 2018, only three months apart.

This happened without the pilots being able to do anything about it, for the reason that they had not been briefed on the system's functioning in the first case and misinformed in the second. In addition to the traditional ethically problematic aspects of the issue - information and training on overriding the system were optional functions that were paid for by the companies involved - there is the fundamental problem of assuming pilot incompetence versus system competence. It's as if we presuppose that it is so obvious that on-board electronic systems are infinitely more *"*intelligent*"* than humans, that we no longer even inform humans - in this case, pilots – of what systems do and how they do it as they take the place of humans. There is not only an ethical problem here, but also a political problem. It's not the ethics of the so-called "intelligent" systems that are at stake. It's the ethics of their creators, who are human, (all too human, to borrow from Nietzsche), and how they assume they know what is good for others, instead of letting others have a voice. It is not the machines that are at the root of this assumption of human incompetence, to the point of not even informing people what is being implemented and what they are directly impacted by as users. It is the men and women whose training in the ethical and political issues of the systems they manufacture is nonexistent.

We are at the heart of ethics, the continuation of which is, according to Aristotle, politics. To speak plainly, this "ethical" problem that we attribute to what is called "artificial intelligence" plays out on the backdrop of an eternal problem of all political life. It's a power struggle, shown by the fact that some people consider themselves more knowledgeable than others, or even knowledgeable at all, and they treat others – users – like incompetent children. This is the problem of all tyranny, all dictatorships, all oppression. We find ourselves in the exact same problem when we realize that *the product sold by companies like the web giants is the users themselves,* that the robots are supposed to know better than they know themselves. The assumption here is that human behavior is predicted by past behavior: we are only supposed to love what we have always loved. So what meaning does the "future" truly hold?

The difficulty is that the problem is voiced quietly, so to speak, via the supposed objectivity and neutrality of the technologies and the intelligence supposedly presiding over them. Clearly, the ethical problems posed by artificial intelligence have nothing to do with the systems as such. The "ethical" problems linked to artificial intelligence are in turn linked to the idea that creators have of the relationship between humans and non-humans. Here, as everywhere, the biggest difficulty is that the victims of this dynamic are often complicit in the oppression imposed on them or the power exerted over them.

If it is urgent, to correctly ask the question about ethics in artificial intelligence, to read the *Discourse on Voluntary Servitude by Étienne de la Boétie*, it is even more urgent to keep in mind that machines are not responsible for what we make them to or what we dream of making them do. It is us, the humans, that are ultimately responsible for the machines we dream up. The ethical problems of what we call artificial intelligence are in fact the ethical problems linked to the excesses of human imagination. It is therefore Plato's *Republic* that we must read to ask these questions in the best way possible.



© iStock - Peepo

GENDER INEQUALITY: IS AI A BLESSING OR A CURSE?

Estefania Santacreu-Vasut *is Associate Professor of Economics at ESSEC Business School. Her research focuses on gender and institutions and has been published in outlets such as the Journal of Development Economics, Journal of Economic Behavior and Organization, among others. Estefania has advised the EU on its gender equality strategy. She has also been a consultant for the OECD, is a co-founder of the project Gender & Finance, and the co-author, together with Tom Gamble, of the popular press book The nature of goods and the goods of nature: why anti-globalization is not the answer.*

Since 2010, in the post #Metoo world, the causes and consequences of gender inequalities have come under increasing scrutiny from academics, policy-makers, consumers and the general public. Also during the last decade, concerns about the diffusion of artificial intelligence (AI) have attracted increased attention in the public debate. AI is a "general purpose technology (GPT), the advances of which create a drop in prediction costs, especially thanks to the "machine learning" domain (Agrawal, Gans, & Goldfarb, 2019). Meaning the use of data to make predictions. One area that will strongly be impacted by AI is the labor market, a market where gender inequalities have been particularly studied by social scientists. The gender wage gap (the average difference between the wages of men and women) has been deconstructed to investigate the role of attributes (for example differences between men and women in years of education, occupational choices, years of experience...) and the role of discrimination (different effects of the same attributes). Discrimination is often measured as the part of the gap that is left unexplained after controlling for all observable differences between men and women. A difficulty researchers face when measuring it is to make sure that all differences are taken into account, as some of these differences may be hard to measure and not available in the data. Because AI contributes to lower data prediction costs, it is not surprising that the debate surrounding AI has also led to questions regarding the fairness of AI algorithms, or AI decision-making. Will AI algorithms help reduce gender discrimination, for example by improving predictions on workers' productivity based on objective factors? Or, on the contrary, will they exacerbate inequality in hiring and remunerating workers? If we look beyond the labor market, how gender biased is AI? While answering these questions is an ongoing endeavor that will require increasing research resources, three considerations are worth taking into account. First, defining the correct benchmark (or counterfactual). Second, distinguishing between algorithms' objectives and predictions. Third, when formulating policy advice, taking into account the consequences of informational asymmetries between regulators and AI users.

© GettyImages - Gremlin

The role of the counterfactual

Examples about AI exhibiting gender biases have reached the popular press, shaping the public perception that AI leads to discriminatory decisions. Yet this evidence in itself is insufficient to discard AI algorithms. The key question for policy-makers may not be: "are AI algorithms prone to gender bias?" but rather "is the size of such bias bigger or smaller than without using AI algorithms?". Indeed, the alternative to using AI algorithms is to rely on human judgment and decision-making. As extensive research shows, human decisions are often prone to gender biases. In recent work with my colleague Professor François Longin (Longin and Santacreu-Vasut, 2020), we show that this is the case in an investment context, an environment where decision-makers have the objective to maximize their gains and do not explicitly pursue a gender biased objective. Yet investment decisions are prone to unconscious biases and stereotypes that lead to biased trading choices, for example, selling stocks when a female CEO is appointed to lead a company. While this may not be the objective of traders, investors may predict that selling is the best course of action partly as a result of their gender stereotypes.

The distinction between objectives and predictions

The distinction between objectives and predictions is central in economic theory. This distinction is extremely useful to think about the fairness of AI (Cowgill and Tucker, 2020). Are the goals of an AI algorithm biased? Or are its predictions biased? To answer this question, it is important to distinguish between different types of algorithms, in particular, those that are fully automated versus those where a human is "in the loop". Similar to investors in the financial market, programmers or the "human in the loop" may have unconscious biases that translate into biased algorithms even when the goal of the algorithm is unrelated to gender. Programmers may be biased because, like many of us, they may suffer in-group bias (Tajfel, 1970), meaning the tendency of individuals to distinguish between "we" and "them", deeply embedded in our socialization process. As programmers may be mostly male, they may suffer from homophily: the tendency to interact with individuals from their own group, including their same gender. How then should we deal with such biases? Are legal tools beneficial?

Policies to counter biased objectives and biased predictions

Using legal tools may be beneficial to fight gender bias when regulators identify that the objective of an AI algorithm is biased. Yet using stringent legal tools can incentivize programmers and users to create less transparent algorithms, increasing the informational asymmetry between the regulator and the regulated regarding the algorithm's objective. More radically, firms and organizations may decide to avoid using AI algorithms to decrease the scrutiny from its stakeholders as well as from regulators. Developing more transparent algorithms may therefore lead to a trade-off between incentives ex-ante and incentives ex-post.

The policy tools to fight biases in predictions, on the contrary, may need to rely less on legal tools and more on education. For instance, we should educate future decision-makers to undo some of their own biases and to recognize that data used by algorithms may itself contain biases. For the current generations, it is important to develop training programs that tackle the source of gender inequalities, namely human biases. In sum, whether AI will be a blessing or a curse for addressing gender inequalities will depend on fighting the root of gender prejudice: not machines, but humans.

References

A. Agrawal, A., Gans, J., & Goldfarb, A. (2019). Economic policy for artificial intelligence. Innovation Policy and the Economy, 19(1), 139-159.

Cowgill, B. & Tucker, C. E. (2020). Algorithmic Fairness and Economics. Columbia Business School Research Paper, Available at SSRN: https://ssrn.com/abstract=3361280 or http://dx.doi.org/10.2139/ssrn.3361280

Longin, F. & Santacreu-Vasut, E. (2019). Is gender in the pocket of investors? Identifying gender homophily towards CEOs in a lab experiment.Available at SSRN: https://ssrn.com/abstract=3370078 or http://dx.doi.org/10.2139/ssrn.3370078

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American,* 223, 96-102.

© GettyImages - MicroStockHub

THE ETHICAL DILEMMAS OF AI-POWERED GOVERNANCE SYSTEMS

Julien Malaurent *is Associate Professor of Information Systems at ESSEC Business School. He received his PhD in Information Systems from ESSEC Business School and Université Cergy-Pontoise in 2011. Before being an academic, Julien worked for several years in China as ERP consultant for large multinational corporations implementing global information systems. His research works, based on qualitative approaches, address issues related to the use of information systems in multicultural contexts, as well as how organizations digitize their operations. His research has been published in top journals of the discipline such as Journal of Management of Information Systems (JMIS), European Journal of Information Systems (EJIS), Journal of Information technology (JIT). He is also Senior Editor for the Information Systems Journal (ISJ).*

Thomas Huber *is Assistant Professor of Information Systems at ESSEC Business School. He received his PhD in Information Systems from the University of Bern. His research is mainly qualitative and process-oriented and focuses on the dynamics of governance and control in inter-organizational software development, software platform ecosystems, digital collaboration, and the management of digital transformation and AI projects. His research has been published in the Journal of Management Information Systems (JMIS), Information Systems Research (ISR), Information Systems Journal (ISJ), European Journal of Information Systems (EJIS), IEEE Software, and in proceedings of the field's leading conferences.*

Triggered by an event in your calendar, your smartphone assistant reminds you to leave for work early. News updates tailored to your morning habits pop up on your phone. The daily decision to click on them or not further refines the parameters of your readership profile. You arrive at work, respond to emails, order raw materials that are low on stock according to the ERP system, and then head home to take a run. Over the course of the day, you have left a trace of personal data all over: the words you spoke in your own home recorded by your smart speaker, the conditions of your commute recorded by your phone, the nature, content, and performance of your daily work carefully tracked by enterprise systems, how far and where your run takes you precisely measured by your fitness wearables. Until you go to bed, at which point you will feed your health app with your sleep patterns, you will further add to this data trace: the contents of private text messages, the items of your shopping basket, Google search queries, Facebook profile visits, your choice of mates and dates. Your private and professional activities performed in the digital world, all recorded, rendered as data, analyzed, processed, and exploited.

© GettyImages - Prostock-Studio

This may sound like the envisioning of a privacy nightmare fraught with ethical dilemmas – but it is in fact the present. It is the state of the world in the age of what Harvard University's Shoshanna Zuboff calls 'surveillance capitalism'. Yet this age has only just begun. The next wave of large-scale, AI-powered systems is already in the making and will bring new ethical dilemmas and challenges because they touch on the very core of societal vs. individual well-being, algorithmic certainty vs. human agency, and technological determinism vs. moral responsibility. In this article, we briefly present the current status of AI-powered governance systems, and suggest directions to understand how the organizations and individuals that are developing and using these systems could cope with the ethical dilemmas that they may encounter.

The current generation of AI-powered systems such as the services offered by Google or Facebook set out to gather, process, and commodify the totality of information about our every thought, word, and deed expressed in the *digital world.* This entails data such as social media behavior, purchasing habits, and credit card transactions. The next generation of AI-based systems will go far beyond that. As emerging technologies such as context-aware sensors, facial-recognition cameras, and crowd-based reporting regimes become more ubiquitous, there will be an increasingly complete data record not only of our actions performed in the digital but also in the physical world. This data will be on an unprecedented level of granularity and can be used to not only predict individual behavior but also to shape, produce, modify, manipulate, and control it.

Not only the private sector with its power - and resourceful tech giants - will be a key player as a key sponsor of these systems. In fact, many nation states have started to work on large-scale, AI-powered governance systems. The Chinese Social Credit system is a prime example of these AI-powered governance systems. These systems come with the promise of optimizing almost all aspects of a populace ranging from energy consumption, to health and sleep. By increasing the transparency of human behaviors and quantifying even the most personal acts, nation states can entice a competitive arms race of social desirability and the will to improve between their citizens.

Two Ethical Dilemmas of Large-Scale AI-Powered Systems

Despite the many promises of such all-encompassing AI-powered systems, they also raise important ethical questions, not only from the perspective of individual citizens and the wider society but also from the perspective of the designers and developers of AI-powered products and services as well as the managers overseeing design and development. The first ethical dilemma pertains to the gathering and exploitation of data. On the one hand, the creators of AI-powered systems have strong incentives to gather and exploit ever-larger amounts of data. The reason for this insatiable hunger for data is the technical logic of modern AI systems. These systems become more accurate the more data they are fed with. On the other hand, the more the providers of AI systems use this data to shape and control behaviors, the more they restrain human agency. They replace free will and individuality with algorithmic certainty and guaranteed outcomes (Zuboff, 2019). Hence, an important ethical challenge is to balance the wider societal needs that these systems are designed to serve

with the individual need for freedom of expression and free choice.

A second ethical challenge is that of the moral responsibility for the behaviors of the AI-based system. If an AI shows some undesired behavior such as discrimination and biased decision-making, who is responsible for it? Let's stop thinking that machines are bad or evil. Algorithms do not distinguish between "good" and "bad" people or good and bad behaviors by themselves. Those systems have been produced by IT professionals, i.e., by designers, developers, and engineers. These professionals have fed them with classifying and clustering rules or at least with the datasets that they deemed appropriate to extract these rules inductively. Yet one cannot easily assign responsibility to the engineers alone either. The lines for moral responsibility are blurred by the fact that due to their intrinsic nature, artificially-intelligent systems are to some extent autonomous; they learn to show behaviors that go beyond their original programming.

Investigating the Management of Ethical Dilemmas in the Age of AI

ESSEC has positioned itself around three strategic pillars: Enlightening Entrepreneurship, Together, and the Metalab. The first aims to invest in our entrepreneurship and innovation ecosystem. The second is the creation of knowledge that will help the current and the future generations of leaders to act not only with business acumen but also with social responsibility (Together). The third is the creation of knowledge that will help leaders navigate through the "storm" of technology-driven change processes such as the digital transformation (Metalab). We believe that in order to ensure that the next generation of AI-based systems will not only benefit a few private and public organizations but also strengthen the societal institutions that liberal democracies rely upon, the ethical and technical aspects of management need to become two sides of the same coin.

Therefore, we have recently started a research project that takes an in-depth

look at one of the largest AI initiatives in the world. In this research project, we 'unblackbox' the complex interplay between instrumental managerial considerations such as technical proficiency or economic efficiency with ethical value orientations. Our goal is to understand how both ethical and instrumental considerations become entangled with modern AI-powered systems that weave together smartphones, stationary computers, CCTV systems, and other sensory devices to form a technological apparatus of unprecedented complexity.

Through this large-scale investigation, we strive to uncover which ethical dilemmas the actors involved in the development and improvement of these systems encounter and how they cope with these dilemmas. We moreover expect to considerably advance current understanding of how the technical, the managerial, and the ethical intertwine in real-world AI projects and together drive both intended and unintended consequences. Overall, we are hopeful that this research project will produce actionable managerial levers for designing and maintaining large-scale AI-powered systems that are ethics savvy. ■

Note

This article presents the main motivational aspects of a large research project carried out by Prof. Huber and Prof. Malaurent. This research is funded by ESSEC Foundation under the White Project grant won in January 2020. For further details, please contact the authors at: huber@essec.edu and malaurent@essec.edu

Reference

Zuboff, S. (2019) *The age of surveillance capitalism: the fight for the future at the new frontier of power,* Profile Books.

TOWARDS A POLICY OF ALGORITHM SECURITY?

Guillaume Chevillon *is Professor at ESSEC Business School where he is Academic Co-Director both of the Metalab for Data, Technology & Society and the ESSEC|CentraleSupélec Master in Data Sciences & Business Analytics. He researches econometric theory with applications to forecasting and statistical learning in economics and finance. Applications span economics, finance, energy forecasting, social networks, global warming and pandemic dynamics. Guillaume obtained an MPhil and a DPhil in Economics from the University of Oxford and an MSc in Engineering from Ecole des Mines de Paris. He is a member of the OECD network of experts on Artificial Intelligence and has been a visiting scholar or professor at Brown, Oxford, NYU, Keio, UNSW Sydney and the NY Federal Reserve.*

While Elizabeth Warren and Bernie Sanders, both senators and candidates for the US Democratic presidential primaries, advocated for the dismantling of the monopoly of the "Big Tech" companies (the American web giants),[1] during their campaigns, the European Commissioner for Competition Margrethe Vestager has declared several times over the last few years that while she shares the same objectives of protection and freedom of users, the solution of dismantling by antitrust laws did not appear effective to her.[2] She finds it more useful to combine the promotion of competition with regulatory constraints such as, e.g. the General Data Protection Regulation or the recent Digital Markets and Services Acts. Even Mark Zuckerberg has himself at times called for more regulation of social networks, while his Facebook co-founder Chris Hughes goes further and advocates breaking the company down.[3] So what should be done?

To improve control by public authorities, some, like Hannah Fry, a mathematician at the University College of London who published a well-received book on data and algorithms in 2018,[4] suggest establishing a Regulatory Authority for Algorithms, following from the principles of the US Food and Drug Administration. Such a proposal deserves attention because it suggests that algorithms – today mainly those of social networks but tomorrow all those grouped under the term of "Artificial Intelligence" – present a potential for danger that must be evaluated before commercialization.

According to this health analogy, algorithms bring a benefit but can have harmful side effects (confinement in informational bubbles, addictive behaviors, collapse of democratic practices…). This is also the core of the criticism voiced by the Netflix-sponsored documentary, The Social Dilemma, that calls for a radical change of business

1 - https://medium.com/@teamwarren/heres-how-we-can-break-up-big-tech-9ad9e0da324c
2 - https://eeas.europa.eu/delegations/united-states-america/43309/commissioner-margrethe-vestager-press-conference-washington-dc_en
https://www.competitionpolicyinternational.com/eu-vestager-says-breaking-up-facebook-would-be-a-last-resort/
3 - https://www.ft.com/content/0af70c80-5333-11e9-91f9-b6515a54c5b1
https://www.nytimes.com/2019/05/09/opinion/sunday/chris-hughes-facebook-zuckerberg.html
4 - Hannah Fry (2018), Hello World: How to be human in the age of the machine, 2018, London: Penguin.

© GettyImages - Cofotoisme

models away from the "human attention extraction" model. This documentary features the Center for Humane Technology, whose president Tristan Harris wrote in the Financial Times (March 2020)[5] that the key issue is not only the ownership and reselling of data so much as the functioning of algorithms of social network platforms whose aims are to maximize personal engagement at any cost. He calls for the regulation of such platforms as "attention utilities", subject to licensing that ensures they operate in the public interest.

Human attention and social bubbles

In this proposal, an independent agency could analyze algorithms ex ante via a "social impact assessment" and, if appropriate, allow their launch. Hannah Fry and Tristan Harris seem to go beyond the proposals of Bernie Sanders, Elizabeth Warren and Margrethe Vestager of surveillance by a public authority: they request administrative control a priori.

The rationale for such control is clear: algorithmic recommendations are there to make us react quickly, not to present us with all relevant alternatives so we can make an enlightened decision. Indeed, social platforms such as Twitter and Facebook are financed by advertising and their interest is therefore to maximize the depth of their network, the audience of the interventions and the amplitudes of the reactions in order to collect ever more information on the tastes, interests and preferences of their users. This allows them to suggest the most relevant ads to us or, like Netflix, who modifies the covers of films and series based on our reactions, to present us with suggestions that will have the most influence on our behaviors.

In the context of social media, this generates the so-called "Man bites dog" phenomenon according to which the most referenced, and therefore shared, information is not necessarily the most relevant but, the most surprising[6] (reversing, here, the most usual "Dog bites man"). Many online media are thus making themselves known through a race for provocative or surprising "news" (who hasn't seen a headline promising "you won't believe what happened to…") that is ultimately not very informative. In return, social platforms do not gather information on our deep and well-thought interests. The message of The Social Dilemma is that maximizing engagement should not be measured through time spent online or the number of interactions. It should instead contain an appraisal of the quality of the interactions.

For want of such a focus on quality, success in recommendations is currently measured through engagement intensity, i.e., whether we respond to stimuli. Hence algorithms model our preferences and interests, relying on the history of our behaviors, sharing and reading activities to better understand us. It then boils down to predicting our future reactions. These predictions are however flawed in that they rely on the partial information that our history provides, not on the wider range of our potential interests. In doing so, they reduce the diversity of suggestions and

5 - https://www.ft.com/content/abd80d98-595e-11ea-abe5-8e03987b7b20
6 - For an application to Macroeconomics, see Nimark (2014), https://www.aeaweb.org/articles?id=10.1257/aer.104.8.2320

our exposure to ideas that disturb us: acting as reinforcement mechanisms, they can lock us into an information bubble. This constitutes the main criticism made by Tristan Harris and the Center for Humane Technology: the information bubbles induced by social network platforms may lead us apart, foster divisions, and tear the social fabric. The Social Dilemma then foretells the end of democracy, using the example of French Gilets Jaunes (the Yellow Vest movement) who shared information on Facebook and WhatsApp. While it is true that the fake news that resides in these social bubbles render dialogue difficult, historians of social movements might argue that progressive groups often develop their own narratives that differ from the perceptions of established media, for instance in the US fight for civil rights in the 1950-60s, or sexual minorities in the 1970-80s. Arguments based on the Gilets Jaunes are therefore questionable: the issue is about the intensity and widespread prevalence of these bubbles, rather than their mere existence.

National Algorithm Security Agencies?

Creating a National – or European, or International – Algorithm Security Agency, as interesting as it may initially appear, forgets an essential element: algorithms of social platforms are not just scientific objects (mathematical or computer based) that can be assessed a priori, as they must be analyzed in a social sciences context through their middle or long term impacts. Short of presenting the dramatic consequences of Skynet in the Terminator movies, any social algorithm – a recipe to get a result (a behavior) using certain ingredients (stimuli) – automatically escapes its designer's control. This is not just because social platforms are indifferent to studying the consequences of their tools, as participants in The Social Dilemma seem to imply.

Indeed, social network algorithms are just a new methodology to achieve an old goal – the very principle of any public policy, in fact: influencing the behavior of individuals. History is dotted with our failures in this respect. Since the creation of statistical institutes, e.g., the US Census bureau or INSEE in France, and the development of polling techniques, public authorities and private companies have used data and statistics to analyze the behaviors of citizens and consumers: they, in turn, attempt to modify these behaviors to obtain specific results (economic growth, poverty reduction, increased sales...).

However, when it comes to influencing a person, the difficulty lies in that the latter's behavior evolves in response to the influences he or she receives. We, humans, are like machines that change function, shape or mode of operation, as soon as something tries to nudge us on.

Taking human reactions into account

Social and economic sciences have long been studying the reciprocal influences between individuals and their environments, and, in this context, the question of modeling and control. Here, the important question is not whether an algorithm is harmful: as a recipe, it is designed for a specific purpose and generally works reasonably well in the short term. However, it is a partial recipe that only uses a fraction (even if many) of the possible ingredients. When it comes to generating engagement on social networks, because the notions of truth or quality are absent from the current algorithms and only popularity is taken into account, misinformation develops. Thus, the algorithm achieves its short-term goal, but its medium-term impact (polarization of information, lack of contradictory information and prioritization) is not within its purpose. This is the economic notion of externality, whereby companies internalize some benefits but externalize the negative consequences (such as pollution in an industrial context) to society at large.[7]

In social sciences, any stimulus is known to modify not only the individual it affects but also the context in which she operates. Assume that you analyze the behaviors of humans and model them via an algorithm: when you try to use this algorithm to influence people, they find themselves in a new context since someone – now, you – is now trying to modify their "usual" behavior. This, in turn, generates new reactions that can potentially make the algorithm useless or even counterproductive.[8] For example, a famous Internet controversy concerned the use of search history "cookies" by airline companies that were used to identify one's planned vacations to increase the prices of their flights. When this debate first appeared a few years ago, many Internet users played with their flight searches in order to disrupt the cookies and obtain, contrary to algorithmic forecasts, lower prices.[9]

The human dimension is not sufficiently taken into account in algorithms that come from engineering sciences where

7 - Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *American Economic Review,* 75(3), 424-440.
Wattal, S., Racherla, P., & Mandviwalla, M. (2010). Network externalities and technology use: a quantitative analysis of intraorganizational blogs. *Journal of Management Information Systems,* 27(1), 145-174.
Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives,* 31(2), 211-36.

8 - This is an example of the famous Lucas Critique that was introduced in economics by Robert E. Lucas (1995 Economics Nobel Prize recipient): Lucas, R. E. (1976). Econometric policy evaluation: A critique. In Carnegie-Rochester Conference Series on Public Policy 1(1), 19-46).

9 - https://time.com/4899508/flight-search-history-price/
http://www.businessinsider.fr/us/clear-cooking-when-searching-for-flights-online-2015-9

individuals are seen as black boxes: their thoughts are not perceived but their consequences are measured through the resulting actions. In reality, humans think about the influences exerted on them, and they can counteract them.

Ensuring long-term sustainability

In such a context, it seems illusory for an administrative authority to control ex ante the tools of artificial intelligence because their medium-term consequences are almost unpredictable given the number of actors that hold an influence. An adequate answer to the question posed by The Social Dilemma may therefore be found outside the context of drugs and medicines, but closer to that of controlling inflation.

Public authorities have long aimed to avoid the twin pitfalls of inflation that is too high (hyperinflation causing political instability in the 1920s) or too low (the deflation leading to impoverishment in the 1930s). Moderate inflation is optimal, but it is an unstable equilibrium, the result of the individual decisions of millions of individuals and companies, decisions that are themselves the result of people's perceptions of their environment (past, present and future) and the decisions of others (competitors, suppliers, customers...).

After thinking that governments could directly control prices (e.g., the price of bread in France until the 1980s) or policy tools, the consensus in academic circles over the last thirty years has been that the agency in charge of inflation control, the Central Bank, deserves to be independent and in full possession of the relevant tools – not those of direct control of individual prices, but of individual decision making (via interest rates) and supervision of major market operators (banks and financial institutions). In this context, the role of the government is merely to set the objectives of monetary policy (low inflation and, in some countries such as the US, full employment). Central banks were made independent to convince the population of their sole pursuit of mandated medium-term objectives, away from short-term political considerations (that may come into play during election time). We saw an attempt at shifting this status quo when President Trump threatened to replace the Chair of the Federal Reserve to exert an influence over policy tools (interest rates).[10]

A Central Bank of Algorithms

Rather than imposing a priori administrative approval by an Algorithm Security Agency, controlling Artificial Intelligence algorithms may be better achieved by an independent authority that directly supervises AI companies and imposes certain basic algorithms. It could, for example, monitor "essential" algorithms, possess the ability to modify them and obtain daily impact measurements (as the Central Bank checks every night that private banks balance their accounts). This independent agency, this Central Bank of Algorithms, could thus reintroduce a focus on the medium term and the evolution of society in accordance with objectives set by governments. It could also monitor the degree of concentration of current Internet platforms, to avoid the emergence of companies that are "too big to fail" and endanger the whole system.

Its ability to act directly, its independence and its focus on explicit objectives would help foster systemic trust by all individuals and businesses. This confidence is the key factor that makes it possible to better anticipate the reactions of individuals to the stimuli received: it improves reactivity and facilitates the resolution of the key issues posed by information bubbles and misinformation. As with financial innovation that is constrained by regulation to avoid major economic crises (which happen nevertheless when regulation is lowered), the development of artificial intelligence might be slightly slowed but with an objective of public interest and a benefit of long-term sustainability. ■

10 - https://www.newyorker.com/news/our-columnists/the-high-stakes-battle-between-donald-trump-and-the-fed

SO MANY WAYS TO BE WRONG ABOUT THE FUTURE OF AI, BUT IT'LL OUTSMART US IN THE END

Fabrice Cavarretta *is Associate Professor in Leadership and Entrepreneurship at ESSEC Business School. Coordinator of the PhD Entrepreneurship seminar, he developed various intrapreneurship programs through ongoing learning. His research focuses on logic used by managers to develop new firms, and on artificial intelligence applied to people and organizations analysis. Profesor Cavarretta recently published Oui ! La France est un paradis pour les entrepreneurs (Plon, 2016) in which he tackles the specificity of national entrepreneurial ecosystems.*

In 2015, I began to examine how artificial intelligence could be used in management. AI was becoming popular again, so it was high time to understand if - and how - the tools had evolved. Could AI be used in ways other than in well-defined cases like image recognition? Had it begun to have the flexibility necessary for discovering complex relationships in complex data, like those between individuals, teams, performance, and emotions in a company?

After a few years of investigation, I decided to postpone this project, and to prepare for a new winter of artificial intelligence, a time of slow progress, during which we wonder if the technology will ever truly take off.

The current AI algorithms are undeniably impressive. Their ability to respond to predetermined questions is extraordinary: for example, they can now identify cats, individuals, and cars, in increasingly natural situations.

But for each new use of AI, there are even more cases where obstacles hinder its practical usage: data that is unavailable, not properly formatted, or too "noisy"; an absence of analytical ability and trained personnel; ill-adapted technology; legal or ethical issues, etc. The list of good reasons for not using AI, even in a narrow domain, could be a research subject in and of itself.

Paradoxically, even in the face of the current multiple obstacles, AI will likely reach its "Holy Grail", and become indistinguishable from human intelligence.

Can AI be intelligent in the human sense?

The current practical obstacles are small compared to the immense difficulty of reproducing typical human cognitive processes. For example, neither intuition (cognitions emerging out of the historical accumulation of experiences) - nor emotions (which provide autonomy through "motivated cognitions") - are currently properly simulated in AI systems.

If AI can address specific questions in specific domains with well-defined data, does it have the ability to find useful relationships in heterogeneous and complex dataset? It is in this capability to produce a logic in a vague, yet motivated,

© GettyImages - PhonlamaiPhoto

way that humans still set themselves apart from current machines.

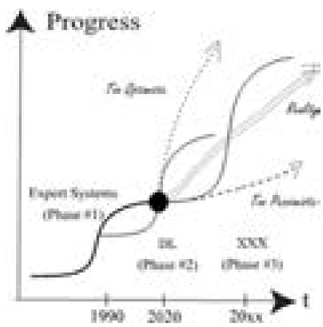
Experts differ in their visions of the future: some talk about AI as a sure thing that is on the verge of taking over the world and surpassing humans. Others swear that machines will never display intelligence in the same way that humans do. Who to believe? Does our world overestimate or underestimate the speed at which AI will progress toward general intelligence?

AI: a narrow scope in the short term, a broad scope after future updates…

Paradoxically, the answer is… "both." This paradox comes from a classic innovation phenomenon. When a field evolves, change does not happen in a linear, gradual way. The response of the human, social, and technical community to an evolution follows S-curves: at the beginning, progress is slow, then there is an explosive uptake, then it slows down again. New evolutions accumulate over time, and combine in a way that can be hard to interpret. Take a look at a visual depiction of this in the figure below, where the three S-curves represent three successive waves of innovation.

Different predictions are possible depending on what we look at. If we focus on the first wave of innovation (#1), which is no longer progressing, we might conclude that technology will not amount to much - an overly pessimistic prediction. Paradoxically, we will also be too pessimistic if we look at the budding wave #3, which has not yet progressed very far.

On the other hand, if we project the speed of the current wave (#2), we overestimate progress and project the future overly optimistically.

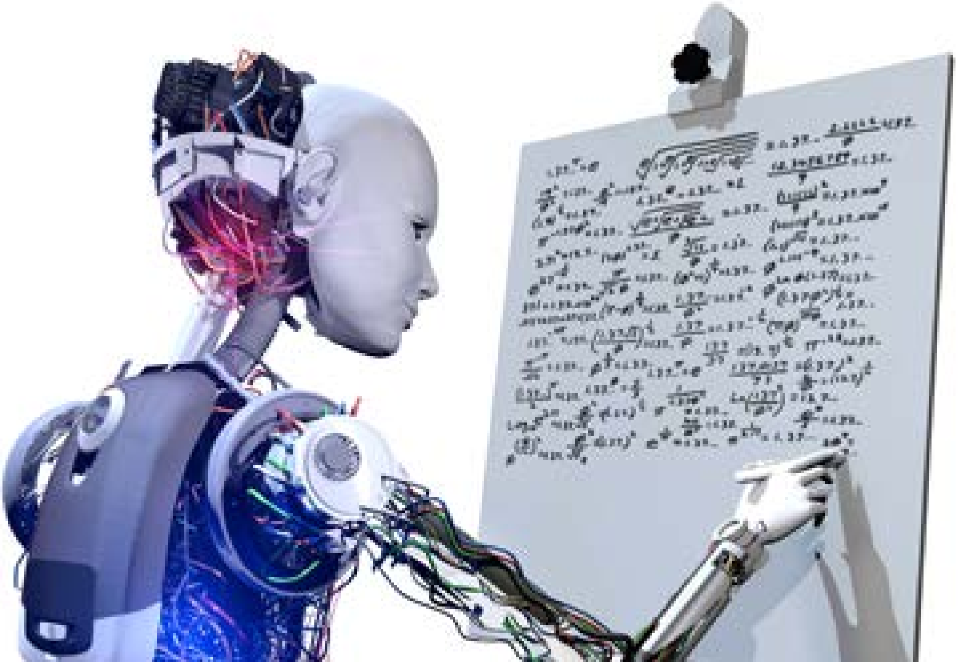


This logic is inaccurate: in reality, the progress from the second wave will calm down, but the progress anticipated from wave #3 will indeed occur, and will likely be followed by progress to come in future waves #4, #5, and #6. It is this succession of innovations, including those that are currently unimaginable but still to come, that forms the real trajectory. Paradoxically, the real trajectory is somewhere in the middle and hard to imagine, because *it does not fit any of the current trends!*

AI in progress

Imagining the future of AI is difficult and troubled by contradictory perceptions about progress. Let's take a step back and remember that AI belongs to the field of computer science, a field that developed over several centuries.

The idea of the calculator can be traced back to the Age of Enlightenment, but at that point, it was thought that the invention would use mechanical machines and therefore be very limited. In the mid-20th century, electromechanical developments and then the first generation of electronics changed the scale of technology, and the newfound power of calculators gave them practical applications in the workplace (like the emergence of IBM). At the end of the 20th century, the miniaturization of semiconductors led to a new wave of progress and



© GettyImages - Devrimb

widespread usage (like the emergence of mass technology, e.g. the PC). Finally, from the dawn of the 21st century, the widespread availability of machines and networks changed the game again, with the Internet wave, mobile devices, and "smart" objects.

At each step, technology drew a lot of pessimism ("but machines are very limited!") and a lot of optimism ("machines will take over the world!"). So, for AI technology, where are we at? We are in the middle of the road, in the uncertain position of having only observed two waves of progress thus far.

The first wave, starting in the 50s and going full speed in the 80s and 90s, assumed that machines would prove their intelligence in a "symbolic" way. AI was modeled as the work of an engineer manipulating concepts through logic. For example, in expert systems controlling nuclear power plants, AI took the form of a system of rules of the form: "if (alarm = triggered & valve = open) then (close valve)". This was very exciting and gave rise to a lot of studies and some successful implementations.

Unfortunately, this led to an impasse, as it required humans to program it all explicitly, in a circularity that lacked growth potential. The 2000s brought about a first winter for AI specialists, a phase in which there was not much hope of drastic future progress.

At the beginning of the 2010s, the situation was turned upside down: after decades of obscure studies, researchers working on computer vision produced impressive results relying on statistical models.

Paradoxically, this had not initially been considered as AI since, at the time, AI was considered as symbolic manipulations. Their method was dubbed "deep learning" (DL), as it was based on a large number of virtual neurons. With deep learning, machines can execute relatively sophisticated tasks, like recognizing a cat in the millions of pixels of an image.

Above all, these techniques allow for relatively automatic learning, called "Machine Learning", as long as we provide the machine with a massive dataset and the human designates what to search for. For example, given a large number of photos tagged as having a cat or not, we can *automatically* train an algorithm to "spot cats" in photos in general.

We are in the middle of this second generation AI boom, and the results are impressive. If we focus only on the recent successes of Machine Learning, we might conclude that the sky's the limit, that AI is going to take over the world … now! This is an illusion, and corresponds to the "too optimistic" projection in the figure, overestimating the future.

In fact, Machine Learning's strength lies in its simplicity, hence the criticism that this type of AI is nothing more than sophisticated statistics. Slightly insulting, this suggests that statistics can only be a method by which a hypothesis made by a human can be validated in data. ML is indeed mainly statistics, because humans are always supposed to play the central role in imagining the relationships between the entities of the world, and the machine performs them automatically. Note that the machine has not "imagined" anything yet in this approach…

Current ML is limited to predefined tasks: we cannot assign it more complex tasks like automatically cleaning and choosing data, detecting relationships between variables, or identifying variables that might be of interest. Above all, this AI does not suggest explanatory mechanisms: for example, determining that "the photo contains a cat because it is an animal with pointed ears and a mustache". Not only does this technology not suggest anything, but it is also not able to tell us clearly what cues were used because current algorithms are not yet designed for such explication.

Regarding all these questions, the slow progress in current technology suggests that we will never succeed, that human-like intelligence is too complicated, that our expectations are too high. This, too, is an illusion, illustrated by the "pessimistic projection" in our figure, which underestimates the future.

The real trajectory, like any uncertain prediction, is difficult to perceive, landing somewhere between those two trends. AI will have to go through many waves of progress to eventually reach a human-like form of intelligence. This progress will most likely take place, but in the meantime, we will probably go through other "AI winters", the phases during which things move slowly and when no one has faith that the future is bright.

The path toward a general artificial intelligence

At the moment, ML technologies have a single layer: the human specifies the inputs and uses the results. Sometimes an engineer may manually decide to take a result from a first algorithm and input it as the input into a second algorithm. But in the future, there is nothing to prevent this looping from being put into the ML process, and that the relationships will be automatically chained together much like in the human brain. Here are the updates to expect for this to materialize.

The first step forward is to take into account the symbolic in current machine learning techniques. At the moment, ML is efficient for processing a lot of rather continuous data (sounds, images) in order to guess a pattern (distinguish images with cat from those without cat). Unfortunately, for the time being, ML does not work on such symbolic data, meaning discontinuous data in smaller quantities.

The second important advance is the ability to identify causal relationships. Current algorithms are unable to identify, organize and test a logical system based on the data, and are unable to build causal inferences alone. Inference is defined as guessing which factor (e.g., gender, education, etc.) influences which other factors (e.g., salary, promotion, etc.). Current AI can help confirm such relationships, without being autonomous at imagining and proving causality.

The first two important advances to be expected from ML are therefore to be able to automatically detect all categorizations in the data (symbols) and to start linking the symbols together, in a causal way if possible. Once it is able to automatically detect all categorizations in the data (symbols) and to start linking the symbols together, it will then be possible to manipulate them with the old techniques developed in expert systems. These allow recursion, the ability to make inferences about inferences, i.e. to reason about reasoning. The third fundamental advance in AI will therefore be the combination of expert systems with ML techniques, combining the symbolism and recursion of these old techniques with the massive computational scales permitted by the new ML techniques.

Winter is coming while we wait for machines to become motivated

While we wait for these to occur, only humans know to build knowledge, i.e., to pick what to analyze, to form hypotheses, to check them, etc. At best, humans can use current AI technology as a sophisticated statistical helper.

The technical progress required for AI to be able to contribute relatively autonomously to the knowledge development process is astounding. First of all, it necessitates phenomenal volumes of computational power compared to current capabilities. For comparison, human brains are several orders of magnitude more efficient than silicon, both in computational power and in energy consumption. But Moore's law has never failed so far, and it is therefore a safe bet that computational capacity will continue to rise in an ever-increasing and astonishing way.

Even more critical is the fact that these recursive calculations must be conducted continuously permanently, hence all cannot be carried out "ad infinitum" on all items. It will therefore be necessary to invent a computer science based on learning tradeoffs: not only will the machine decide on its own to initiate a search for inferences, but it must also know how to stop and be satisfied with a good enough model. It must also know when to resume learning when necessary.

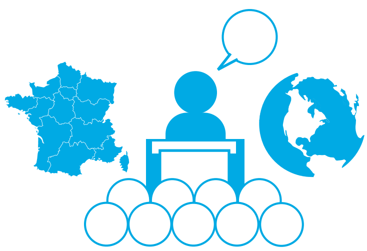
This process could look a bit like what we call "motivation" in human cognition. Essentially, humans are constantly learning about their environment, motivation being a crucial mechanism in the choice between Learning vs. Acting. Machines, on the other hand, are simplistic in that the expert operating them currently decides when and how to do calculations. As with mankind, intelligence will only appear when machines exhibit a form of free will. So far, the modeling of these forms of emergence, of motivated cognition, has not really begun.

All in all, there is no reason not to imagine that AI will become much more flexible and user-friendly than today's limited algorithms. Nevertheless, it is likely that it will take a long time, at least one long winter, or even several successive winters, before reaching that mythical day when the machine is as intelligent as a human. Astonishingly, this artificial intelligence, like human intelligence, will only emerge out of strong motivational mechanisms. ■

RESEARCH SINCE 2015

557
RESEARCH ARTICLES AND BOOKS

59
RESEARCH AWARDS RECEIVED



816
ARTICLES PRESENTED AT ACADEMIC CONFERENCES (IN FRANCE AND INTERNATIONALLY)

ESSEC
BUSINESS SCHOOL

ESSEC Business School
3 avenue Bernard-Hirsch
CS 50105 Cergy
95021 Cergy-Pontoise Cedex
France
Tel. +33 (0)1 34 43 30 00
www.essec.edu

ESSEC Executive Education
CNIT BP 230
92053 Paris-La Défense
France
Tel. +33 (0)1 46 92 49 00
www.executive-education.essec.edu

ESSEC Asia-Pacific
5 Nepal Park
Singapore 139408
Tel. +65 6884 9780
www.essec.edu/asia

ESSEC Africa
Plage des Nations - Golf City
Route de Kénitra
Sidi Bouknadel (Rabat-Salé)
Morocco
Tel. +212 (0)5 37 82 40 00
www.essec.edu

Press Relations
Vitaline GOMES
Tel. +33 (0)1 34 43 30 29

Editor-in-chief
Julia SMITH
Tel. + 33 (0)1 34 43 98 25