INFORMATION SYSTEMS, WEB AND PERVASIVE COMPUTING SERIES

# Automated Data Analytics

## *Combining Human Creativity and AI Power using ChatGPT*

**Soraya Sedkaoui**

iSTE WILEY

# Table of Contents

# List of Tables

# List of Illustrations

*This book is dedicated to my little family, who light my way. May it give back to you even a fraction of what you bring me.*

*With all my gratitude and affection.*

*Soraya Sedkaoui*

*Series Editor*

*Jean-Charles Pomerol*

# Automated Data Analytics

## *Combining Human Creativity and AI Power using ChatGPT*

Soraya Sedkaoui

ISTE WILEY

First published 2024 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.



ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com



Library of Congress Control Number: 2024941677

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN 978-1-78630-978-5

# Preface

> Comparing the capacity of computers to the capacity of the human brain, I've often wondered, where does our success come from? The answer is synthesis, the ability to combine creativity and calculation… into a whole that is much greater than the sum of its parts.
>
> *How Life Imitates Chess,* Garry Kasparov (2007)

Data analytics is a crucial process in today's data-driven world. It involves collecting, cleaning, transforming and analyzing data to uncover useful information, insights, trends and patterns which inform business strategy, decision-making and process optimization. Traditionally, data analytics was a manual process requiring data scientists and analysts to prepare and process data before analyzing them. This was both tedious and time-consuming. The advent of machine learning and artificial intelligence (AI) has transformed data analytics by automating some parts of the process.

Generative AI models, like ChatGPT, are at the forefront of this automation revolution in data analytics. These large language models can understand human prompts and generate coherent and human-like textual responses. They are trained with massive text datasets, enabling them to perform a variety of language-based tasks. Generative models, such as ChatGPT, can be fine-tuned for specific applications, including data analytics.

We can consider these generative AI models as children with unlimited potential. The data scientist's role is to nurture these models, train them and help them grow – just as parents do with their children. In the beginning, these

models are like children – they have immense capabilities, but need guidance to realize their potential.

The data scientist trains them step-by-step, teaching them the different tasks, operations and functionalities required for data analytics. This includes data preprocessing, cleaning, feature engineering, modeling, evaluation and explanation. Models are trained thoroughly, across diverse datasets, learning the nuances of each analytical task.

Gradually, just as a toddler learns to walk, talk and eat solid foods, generative models become capable with regard to various data analytics workflows. With each iteration, their skills improve: they learn to handle diverse datasets, manage missing values, transform features, select optimal models, critically assess performance and generate data-driven insights.

After extensive training on a wide range of datasets and analytical tasks, these generative models evolve from toddlers into mature analytical assistants. They move from simple memorization of problem-solving techniques to the development of true conceptual understanding. The models understand why particular data transformations, models and evaluations are appropriate for given scenarios.

In a way, they develop inductive reasoning and deductive logic, just like humans. They understand the principles and evidence-based principles that underlie data analytics workflows, rather than simply memorizing mechanical instructions. This conceptual understanding is what distinguishes generative AI from prior rule-based expert systems.

Thus, when a data scientist prompts a mature, well-trained model like ChatGPT to perform analysis, it deeply understands the request, rather than simply matching keywords. It draws on its conceptual knowledge to analyze

the dataset, select optimal techniques, generate insights and explain the reasoning behind them. And it does so at superhuman speeds, leveraging the computing power of AI.

But does that make these generative models smarter than humans? The answer is no. At least not yet. Although they can outperform humans on narrow tasks within their training domain, these AI models lack generalized intelligence. Human cognition crucially remains far more advanced.

Unlike generative models, humans possess common sense, intuition, imagination, social intelligence, sensitivity and generalized reasoning capacities. We can creatively solve new and open problems that intersect several domains. Humans also have better judgment, wisdom and morality, which temper our technical capabilities with ethics and responsibility.

So, while ChatGPT can analyze datasets and quickly generate information, it lacks the general critical thinking skills to deeply understand implications and assess ethics. Its intelligence is circumscribed by its training data and purpose. It cannot reason its way through completely new scenarios, as humans can through transfer learning.

That said, narrow AI models offer complementary advantages to human intelligence. Their prodigious memory and computational speed enable exhaustive data analytics. Their lack of bias and fatigue ensure consistent performance. In this way, they endow humans with superhuman data processing capabilities.

Rather than competing with AI, we can collaborate with it – combining human wisdom and ethics with AI’s productivity and precision. Together, we evolve data analytics and make it more insightful and responsible. But humans must

remain in the loop to provide guidance, assess implications and ensure alignment with ethics.

An ideal symbiosis is one where humans manage creative and strategic tasks requiring reason, ethics and imagination, while AI accelerates repetitive analytical tasks requiring memory, computation and precision. Similar to Iron Man deploying the AI assistant JARVIS to enhance his human capabilities.

So, while the gap between human intelligence and AI persists, narrow AI models like ChatGPT are still in their infancy. Their capabilities will continue to grow exponentially thanks to increased data size, computing power and algorithmic advances. One day, they may even cross the threshold into artificial general intelligence (AGI).

But for the time being, generative AI enhances rather than replaces humans when it comes to data analytics process. It takes care of the tedious parts, allowing data scientists to focus on creative, high-added-value tasks. It is becoming an indispensable analytical assistant that continues to learn – like a child who grows into an adult over the years with careful attention.

The key is for humans to guide the development of these generative models in a thoughtful and ethical way. We need to focus on beneficial objectives and monitor for harmful biases or abuse. With judicious care and training, AI can usher in an era of augmented analytics – where human and machine intelligences meet and converge for more powerful, yet ethical, data insights. But the human must remain the supervising parent for the AI child.

Rather than wondering when AI will surpass human intelligence, we should focus on how to cultivate beneficial and ethical AI applications today. Generative models like ChatGPT are impressionable children that will shape the

future based on the guidance they receive. Data scientists have a profound opportunity and responsibility: to educate these AI “children”, so that they become responsible and collaborative allies instead of impenetrable adversaries.

Just as teaching helps humans to consolidate our own knowledge, the training of AI models requires us to thoroughly evaluate our hypotheses, biases and best practices. AI development is as much about advancing our own intelligence – codifying disciplines into coherent frameworks, evidence-based principles and methodologies.

Collective training of generative models advances human knowledge across domains. This requires distilling nebulous issues into structured frameworks; formalizing messy tasks into step-by-step workflows; and crystallizing weakly defined domains into rigorous first principles. Teaching AI models through examples helps us to better evaluate solutions, generalize insights and formalize ethics for humans as well.

The future of data analytics is humans and AI working together – combining the human’s imaginative definition of problems, ethics and strategic judgment with AI’s vast memory, exhaustive computation and high-velocity analytical workflows. Neither can match the synergistic value of the two intelligences combined. Data science augments both human and artificial intelligence.

The time has come to actively train this child prodigy: ChatGPT! It holds enormous potential to enhance human capabilities if nurtured properly. We need to nurture it carefully – teaching analytical skills while emphasizing ethics, exposing ChatGPT to a variety of data and scenarios under supervision, so that it moves from mechanical regurgitation to contextual comprehension.

So, let us therefore guide these generative models with wisdom and kindness. Let us instill analytical techniques with values and ethics, guiding them from innocence to maturity. And let us develop artificial intelligence that makes individuals responsible instead of replacing them. Models such as ChatGPT are still in their maturing stages. With the careful supervision of researchers concerned with data ethics, they could evolve into assistants, opening up new horizons of ethically guided discovery, and collaborating more fruitfully with human than they could achieve on their own.

Just as Garry Kasparov's quote states, human success arises from our ability to synthesize the creative and intuitive aspects of cognition with the computational and analytical aspects. When we combine these complementary modes of thinking and reasoning, the result is an emergent intelligence that simply exceeds the linear sum of creativity and calculation. There are synergies and amplifying effects that result from the fusion of different thinking styles, which makes us unique as human beings.

Synthesis creates something with expanded potential and capabilities beyond what creativity or calculation could achieve individually. It is this integration that allows us to shine. This is what Kasparov believes the human brain excels at, compared to computers.

Kasparov's quote eloquently captures the spirit of the collaboration between humans and AI that we advocate for in this book – combining complementary strengths for amplified potential. Just as the synthesis of creativity and computation expands human cognition, the integration of human ingenuity with the analytical power of AI opens up new frontiers in data science.

When we design responsible workflows that harness both modes of intelligence, the results can far exceed their

individual contributions. A capacity emerges from the thoughtful combination of human ethics and supervision with artificial productivity and rigor. This book provides frameworks for harnessing these synergies to advance data analytics. The future holds exciting possibilities as we forge partnerships between hearts and minds.

July 2024

# Introduction

Imagine this: a seamless and interactive experience where you communicate effortlessly with your data, unravel their mysteries and uncover hidden insights through an engaging conversation. This vision can become a reality thanks to recent advances in artificial intelligence (AI) in data analytics.

The data-driven world is increasingly turning to AI to speed up and improve analysis. More specifically generative AI models, such as ChatGPT, are automating the process of interacting with data to discover insights. These generative models have immense potential as analytical assistants if developed responsibly under human guidance.

Just like a curious child who continually asks questions, generative AI enables fluid and intuitive exploration of datasets without the constraints of predefined queries or static reports. Users can engage in an ad hoc and open-ended dialogue, where the path is guided by emerging insights rather than being limited to pre-planned paths.

This creative symbiosis between humans and machine remarkably amplifies the analytical process. Domain experts provide strategic thinking, intuition, monitoring and ethical judgment. Meanwhile, intelligent AI agents handle the difficult computational tasks – rapid analysis of plain-language questions, analyzing huge datasets and generating interactive responses in natural language.

For example, prompted by a human analyst to evaluate sales patterns, ChatGPT could respond with high-level trends, correlations and high-level hypotheses it has identified by rapidly analyzing sales data. The analyst can then interpret these insights critically, ask follow-up

questions to validate hypotheses and direct the AI to explore new perspectives.

This collaborative analysis is far more exploratory and multidimensional than static queries or predefined reports. It combines the framing and monitoring of human imagination with the exhaustive computational power of AI for an exhaustive and open exploration. Insights emerge more organically, guiding data analysts towards unexpected paths and serendipitous discoveries.

Generative AI augments the human analyst by automatically aggregating, processing and visualizing volumes of data exponentially faster than manual analysis. But it is human creativity that sparks new threads of inquiry, asks “why” and “how”, and connects analytical dots to extract meaning and implications.

The AI assistant finds signals buried in the massive noise. The human provides the contextual framework for interpreting signals into meaningful information. Together, they can unravel complex phenomena by intimately engaging with the raw data through a natural conversational flow.

This fluid and unconstrained approach enables data scientists to seamlessly traverse granular detail and broad trends as required. They can quickly zoom in on micro-level data points if something appears anomalous, then zoom back out to visualize macro-trends – spotting unusual events within wider contexts.

A key advantage is that generative AI allows humans to guide the analysis intuitively using natural language, without the constraints of structured queries or predefined analyses. No need to know programming languages or database schema. Users can engage in conversation as the process evolves.

This makes data exploration more accessible to a wider audience beyond just data scientists. Business leaders, frontline staff and others can participate, encouraging diverse analytical perspectives. The AI assistant becomes the great equalizer, enabling more stakeholders to unlock insights from data via a natural conversation.

Of course, like any child learning a skill, an AI assistant requires extensive training under human supervision to become proficient. It must ingest a myriad of datasets and scenarios to move from the mechanical analysis of queries to a true understanding of the principles and trade-offs involved when interacting with data.

Data scientists must teach these conversational models the nuances of analytical workflows – how to manage ambiguity, validate hypotheses, identify limitations, avoid bias, present information responsibly, etc. Ethics and responsibility must be emphasized, as models progress from innocence to maturity.

But once properly developed, generative AI assistants, like ChatGPT, can automate the laborious aspects of data analytics, while humans focus on high-level creative supervision. This enables exponential scaling of data exploration while maintaining a human-led direction. With sufficient guardrails, AI can even suggest new analytical paths for humans to evaluate.

Of course, generative AI has its limitations. Unlike humans, these models lack any real semantic understanding, beyond pattern recognition in training data. They also lack human judgment, intuition and ethics. Unconstrained automation could lead to misleading insights or generalized bias.

As a result, human guidance is crucial when deploying generative AI. Data scientists must assess the ethics, assumptions and blind spots behind any insights generated

by AI. Generative conversational models are still, ultimately, narrow AIs, contrary to generalized human cognition. They excel only in their trained domain.

The ideal symbiosis is rapidly prompted, human-supervised, generative AI mechanical analysis. But humans provide the global framing, subjective judgment and moral responsibility for interpreting, validating and acting on ideas, both critically and ethically.

Neither humans nor machines alone can match the amplified intelligence of their collaboration. Together, they can mine both granular detail and big-picture meaning to uncover hidden insights in massive and complex data. AI expands the scope of analysis exponentially, while human creativity and ethics anchor it responsibly.

Generative AI promises to take data exploration to new frontiers by enabling a natural, fluid form of interactive analysis. But careful design and monitoring are essential if such automation is to be developed ethically and responsibly. When combined with human creativity and ethics, generative AI can open up new horizons of amplified analytical intelligence and accelerated discovery.

As data grows exponentially, AI will become an indispensable partner to augment human analysts. These models are still maturing children. Under the tutelage of responsible data scientists, they could become trusted analytical allies operating under the compass of human values to extract the maximum value from data for the benefit of society.

## Why this book?

The emergence of AI has given rise to visions of a future where analytical tasks are fully automated by intelligent algorithms far exceeding human capabilities. This narrative

evokes both excitement at the possibilities and apprehension at the implications of handing over decision-making power to AI systems devoid of human values. It imagines a world where generative models like ChatGPT rapidly displace humans' footing in data science.

But this dystopian vision fails to take into account the enduring nature of human ingenuity and ignores the narrowness of contemporary AI technologies. As analytical automation is indeed accelerating, thanks to models like ChatGPT, they remain beneficial enhancements to human intelligence, more than adversaries. When developed responsibly, AI systems can amplify human potential to unlock new frontiers in data-driven insights and value creation.

However, careless implementation, focused solely on efficiency, risks undermining human action and responsibility. Maximizing the transformative power of AI-driven analytics requires respect for human ethics and surveillance throughout the process. This balance remains crucial, but precarious in the age of automation.

That is why this timely book, *Automated Data Analytics: Combining Human Creativity and AI Power using ChatGPT*, serves as an indispensable guide to the modern analytical frontier. It charts a prudent course, one where human imagination and ethics harness automation to elevate, rather than overwhelm, human capabilities. The book advocates thoughtful collaboration with AI, and not abandonment.

And this collaborative approach is, indeed, the wisest course. Because at their core, contemporary AI models like ChatGPT remain narrow and limited by the patterns contained in their training data. Although they can perform defined tasks at superhuman speeds, they lack generalized intelligence capabilities.

What we need is a harmonious fusion of the forces of human creativity, ethics and surveillance with the relentless analytical excellence of AI on vast datasets. Neither humans nor machines alone can match the amplified intelligence unleashed by their symbiosis. The book provides frameworks for productively structuring this collaboration for maximum collective benefit.

By shedding light on the inner workings, development processes and inherent limitations of AI systems, the book enables smoother integration into analytical workflows. Laying bare their technical realities helps to delimit optimal roles that emphasize the specialized strengths of humans versus machines. We can design complementary human–AI partnerships by assigning appropriate tasks based on contextual capabilities.

Humans must take the lead in framing problems, interpreting solutions and thinking globally, while AI undertakes data processing, calculation and in-depth analysis. This book discusses various interaction models that enable a continuous flow between human creativity and the machine’s productivity to constantly improve the discovery process.

It highlights techniques to better expose the model’s decision logic, in order to increase trust and accountability. Continuous surveillance of model behavior and fail-safe guardrails are emphasized to ensure safety and prevent harmful abuse. The book also strongly advocates for maintaining sustainable human values and ethics at every stage of analytical automation to keep progress aligned with social benefits.

It explores frameworks for adapting to society and developing specialized human capabilities that unlock unique synergies with AI. All this is geared towards developing AI-driven analytics in a responsible way for

collective human elevation. The book envisions an augmented intelligence, wherein humans and machines elevate each other, not AI supremacy over humanity.

Mainly, it lays the foundations for the ethical integration of automation into the analytical workflow – from strategic problem definition where humans lead, to iterative collaboration where both participate seamlessly, to evaluative surveillance where human judgment remains essential. This makes it possible to benefit from productivity gains without giving up creativity or responsibility.

The book comes at an opportune moment when AI is transforming analysis, but its capacities remain nascent and impressionable, like a young child. It underlines our responsibility as developers, leaders and citizens to wisely guide AI’s progress from these formative stages to broader social benefit. The choices we make today will determine the long-term trajectory of artificial analytical intelligence.

And the book equips us to make these choices wisely by demystifying AI, highlighting its strengths and weaknesses, and providing ethically based normative advice. Whether you are excited or anxious about the revolution of automation, this book is an indispensable decision-making tool for navigating the future analytically.

Its frameworks, examples and principles enable us to integrate analysis responsibly with a human-centered perspective. You will get comprehensive, balanced clarity on collaborative opportunities that elevate without undermining human agency. And it champions respect for ethics as a guiding compass in the development and application of analytical AI.

This book inspires a future where AI makes data insights available to the whole of society for the common

enrichment of human life. Its sound advice steers analytical automation in a prudent direction that amplifies human potential exponentially while keeping humanity firmly in control. The possibilities are breathtaking if we move forward guided by the human values and spirit of collaboration that this book seeks to inspire.

HÉNO Raphaële, CHANDELIER Laure
*3D Modeling of Buildings: Outstanding Sites*

KEMBELLEC Gérald, CHARTRON Ghislaine, SALEH Imad
*Recommender Systems*

MATHIAN Hélène, SANDERS Lena
*Spatio-temporal Approaches: Geographic Objects and Change Process*

PLANTIN Jean-Christophe
*Participatory Mapping*

VENTRE Daniel
*Chinese Cybersecurity and Defense*

## 2013

BERNIK Igor
*Cybercrime and Cyberwarfare*

CAPET Philippe, DELAVALLADE Thomas
*Information Evaluation*

LEBRATY Jean-Fabrice, LOBRE-LEBRATY Katia
*Crowdsourcing: One Step Beyond*

SALLABERRY Christian
*Geographical Information Retrieval in Textual Corpora*

## 2012

BUCHER Bénédicte, LE BER Florence
*Innovative Software Development in GIS*

GAUSSIER Eric, YVON François
*Textual Information Access*

STOCKINGER Peter
*Audiovisual Archives: Digital Text and Discourse Analysis*

VENTRE Daniel
*Cyber Conflict*

## 2011

BANOS Arnaud, THÉVENIN Thomas
*Geographical Information and Urban Transport Systems*

DAUPHINÉ André
*Fractal Geography*

LEMBERGER Pirmin, MOREL Mederic
*Managing Complexity of Information Systems*

STOCKINGER Peter
*Introduction to Audiovisual Archives*

STOCKINGER Peter
*Digital Audiovisual Archives*

VENTRE Daniel
*Cyberwar and Information Warfare*

## 2010

BONNET Pierre
*Enterprise Data Governance*

BRUNET Roger
*Sustainable Geography*

CARREGA Pierre
*Geographical Information and Climatology*

CAUVIN Colette, ESCOBAR Francisco, SERRADJ Aziz
*Thematic Cartography – 3-volume series*
*Thematic Cartography and Transformations – Volume 1*
*Cartography and the Impact of the Quantitative Revolution – Volume 2*
*New Approaches in Thematic Cartography – Volume 3*

LANGLOIS Patrice
*Simulation of Complex Systems in GIS*

MATHIS Philippe
*Graphs and Networks – 2nd edition*

THERIAULT Marius, DES ROSIERS François
*Modeling Urban Dynamics*

## 2009

BONNET Pierre, DETAVERNIER Jean-Michel, VAUQUIER Dominique
*Sustainable IT Architecture: the Progressive Way of Overhauling Information Systems with SOA*

PAPY Fabrice
*Information Science*

RIVARD François, ABOU HARB Georges, MERET Philippe
*The Transverse Information System*

ROCHE Stéphane, CARON Claude
*Organizational Facets of GIS*

## 2008

BRUGNOT Geérard
*Spatial Management of Risks*

FINKE Gerd
*Operations Research and Networks*

GUERMOND Yves
*Modeling Process in Geography*

KANEVSKI Michael
*Advanced Mapping of Environmental Data*

MANOUVRIER Bernard, LAURENT Ménard
*Application Integration: EAI, B2B, BPM and SOA*

PAPY Fabrice
*Digital Libraries*

## 2007

DOBESCH Hartwig, DUMOLARD Pierre, DYRAS Izabela
*Spatial Interpolation for Climate Data*

SANDERS Lena
*Models in Spatial Analysis*

## 2006

CLIQUET Gérard
*Geomarketing*

CORNIOU Jean-Pierre
*Looking Back and Going Forward in IT*

DEVILLERS Rodolphe, JEANSOULIN Robert
*Fundamentals of Spatial Data Quality*