Data Trends
& Predictions

2023

datacamp

Contents

In 2022, we saw significant developments in the field of data. From the emergence of AI-generated content to the growth of low-code data tools and AI assistants—these advancements signal an upcoming paradigm shift, where data-powered tools and machine learning systems will radically transform workflows across various professions.

2022 also saw digital transformation remain a major theme for organizations across industries as they sought to embrace new ways of working, reaching customers, and providing value. As 2023's looming economic uncertainty puts pressure on organizations to maximize ROI from their investments, digital and data transformation will continue to be one of the key levers by which organizations can cut costs and maximize value for their stakeholders.

More broadly, the upcoming year is shaping up to be an inflection point for data and machine learning technologies, one where the impact will be felt across organizations, individuals, and society as a whole. More importantly, the skills agenda will be on full display. As such, in 2023, skills transformation will be the common thread that enables organizations and individuals to lead in this paradigm shift.

Here are the top 10 trends in data, data skills, and machine learning for 2023.

Large language models will transform coding workflows

The natural language generation model GPT-3 was released in 2020, allowing users to automatically generate text by giving a natural language description of the required output. It can be used for tasks like copywriting, summarizing text, and translation.

In 2021, a programming-focused descendant, Codex, was released, allowing users to generate code in more than a dozen languages from a natural language description. Later that year, GitHub Copilot was released. Copilot can suggest the next line of code, complete functions and methods, suites of unit tests, and even complex algorithms based on existing code or comments. This allows programmers to be more productive. As of today, Copilot is widely available to developers across the globe.

Codex's success led to the development of more efficient and versatile code-writing AI and tools. Even with 50 times less data, Google's PaLM can match Codex's performance. Additionally, DeepMind's AlphaCode generates complete programs for competitive programming challenges. Just recently, ChatGPT was released and showcased awe-inspiring examples of code explanation, correction, and generation.

With these many tools releasing or being released, it is perhaps unsurprising that Forrester predicts auto-generated code will comprise about 10% of global code and tests by 2023. While Forrester's prediction refers primarily to software development code, we predict that AI code generation tools will begin to be adopted by the data science and machine learning communities in 2023.

In 2023, we expect AI coding assistants to become further elegantly embedded in more aspects of the software development lifecycle — Forrester's 2023 AI predictions

Read more here

2 Generative AI will transform content creation and generation

In 2022, the release of powerful generative AI tools such as StableDiffusion, DALLE-2, and ChatGPT sparked significant interest in the potential of these technologies to streamline the workflows of content creators across a range of industries.

The growth of generative AI startups has been fuelled by a surge of venture capital funding and the development of advanced generative models. These startups offer a range of applications, from text and speech generation to creating images and videos. Outside of these startups, there are also viable open-source solutions. Stability AI, with its billion-dollar valuation and its flagship product, StableDiffusion, has emerged as a leader in the open-source generative AI space.

Other startups have found success in specific niches, such as copy.ai and Jasper, which offer marketers an easy way to create social media content and website copy. Other examples include Runway and Fliki, which generate realistic videos. As these tools become more widely available, we expect to see increasing adoption of generative AI in 2023, with initial use cases focused on the generation of stock imagery and marketing images, as well as the development of more imaginative applications.

While the potential benefits of generative AI are significant, the technology also raises concerns about spreading misinformation and harmful content. With the ability to quickly and easily generate large amounts of realistic-looking text, images, and videos, generative AI has the potential to be used for malicious purposes, such as the creation of fake news or manipulated media. As such, it will be important for individuals and organizations to remain vigilant and prioritize the development of responsible and ethical uses of generative AI.

Additionally, the use of generative AI in creating art raises copyright issues. Since generative AI algorithms can produce original works based on existing artwork collected in the training data, it's unclear who, if anyone, holds the rights to the resulting creations. This could lead to disputes over ownership and the ability to profit from generative AI art. It will become increasingly important for policymakers and legal experts to address these issues in the coming years and establish clear guidelines for the use of generative AI in the creation of art. Additionally, artists and content creators are likely to continue to find ways of either opting out of having their work included in training datasets or of collaborating with the creators of AI models to have their work accredited and receive compensation. The 2022 partnership between OpenAI and Shutterstock is an example of such a collaboration.

3

AI will fuel productivity increases across a range of professions

The first two predictions touched upon how generative AI has the potential to significantly improve code creation and content creation workflows. While these two use cases are one of the most evident, the advent of strong generative AI will unlock a new ecosystem of applications, akin to how the advent of the smartphone unlocked the potential for apps like Uber and Airbnb.

In the healthcare sector, for example, generative AI can be used to generate large amounts of realistic-looking medical images that can be used to train and validate machine learning models for tasks such as image analysis and diagnosis. This can help improve the accuracy and efficiency of these models, leading to better patient care.

Organizations will also accelerate the use of generative AI to improve their customer support experience through the use of chatbots. Examples of this include the Zendesk Answer Bot or HubSpot Chatbot. Moreover, by combining different generative AI modalities, digital humans could replicate real-life human interaction and transform customer service with life-like digital avatars and voices. Additionally, 2023 could also see the advent of AI tools that can manipulate computers based on natural language instructions. The combinations of these tools can lead to completely reimagined workflows, for example, here is what a reimagined sales workflow would look like:

Before meeting	Sales meeting			Deal closure
Generative AI tools summarizes the prospective customer's position, relevant company challenges that prompted the sales meeting, and other relevant context.	Sales Representative and prospective customers meet to discuss needs. Generative AI tool takes notes.	AI tool uploads prospective customer information into company CRM (e.g., Salesforce), based on outcomes from the meeting.	Generative AI tool sends a follow-up email to prospective customer covering main conversation outcomes, next steps, and selects relevant materials to send.	Generative AI tool takes care of the contract write-up and amendments and is reviewed by the legal commercial team.

DIVE DEEPER INTO GENERATIVE AI

Inside the Generative AI Revolution — Hear from Martin Musiol, Data Science Manager at IBM, on how Generative AI will transform industries.

GPT-3 and our AI-powered future — Hear from Shubham Saboo and Sandra Kublik, authors of GPT-3: Building Innovative NLP Products using Large Language Models, on how GPT-3 will drive an AI-powered future.

A Beginner's Guide to GPT-3 — Learn how to work with GPT-3, and design NLP-powered applications with it in this accessible tutorial.

TOOLS & TECHNOLOGY

The Jupyter Notebook is a critical tool for data scientists, with 66% of data practitioners in the 2022 Kaggle Survey indicating that they use the platform. However, many practitioners have expressed challenges with setting up a Jupyter environment, sharing analyses with colleagues, and accessing powerful IDE tools that accelerate data workflows.

Recent advancements in the notebook interface have addressed these issues, with new interfaces such as Hex, DeepNote, AWS SageMaker, and DataCamp Workspace enabling practitioners to explore and analyze data right in their browsers easily. These interfaces also offer the ability to automate data workflows, such as creating data visualizations, easily accessing GPUs and package tools, version control, and much more.

Organizations will also accelerate the use of generative AI to improve their customer support experience through the use of chatbots. Examples of this include the Zendesk Answer Bot or HubSpot Chatbot. Moreover, by combining different generative AI modalities, digital humans could replicate real-life human interaction and transform customer service with life-like digital avatars and voices. Additionally, 2023 can also see the advent of AI tools that can manipulate computers based on natural language instructions. The combinations of these tools can lead to completely reimagined workflows, for example, here is what a reimagined sales workflow would look like:

How DataCamp Workspace Enables Faster Data Workflows

No installation required	Native SQL Cell Integration	No-code visualization cells
Start coding and analyzing data right in the browser—no setup or installation required.	Native SQL cells make it easier than ever to switch between SQL and Python or R—and connect to your proprietary databases.	No-code visualization cells let you easily create visualizations without coding knowledge. The combination of SQL and no-code chart cells enable any modern analyst to create robust reports.

In 2023, notebooks will win 20% of Excel Users with data apps — Tomasz Tunguz, Venture Capitalist at Redpoint Ventures

Read more here

5 Data observability will drive higher trust in data and accelerate data culture

TOOLS & TECHNOLOGY   DATA CULTURE

Data downtime can be costly, delaying important data-driven decisions and causing disruptions to the work of data engineers. In order to prevent these disruptions and maintain the trust of data users, data observability has become increasingly important. This practice uses best practices from developer operations (DevOps) to provide a complete understanding of the health of an organization's data.

The three pillars of data observability are metrics, logs, and traces, which are used to monitor data quality, track changes, and triage incidents. Tools such as Monte Carlo provide automated monitoring, root cause analysis, and comprehensive data lineage, helping to prevent downtime and improve the overall health of data pipelines.

The productivity benefits of data observability have led to its growing popularity, with most IT professionals agreeing that it would be beneficial for their organization, according to the 2022 State of Observability and Log Management. However, only 11% of respondents have fully implemented mature observability solutions. In 2023, we expect to see more organizations adopting these tools and benefiting from increased confidence in their data and a stronger data culture.

> " If you think about observability in the context of software engineering, it has emerged in the last couple of decades as a fast growing area to support application infrastructure downtime. As such, software engineering teams teams track metrics that help understand the health of their systems. Data teams need to also start looking at data downtime, and apply an observability lens to data quality."
>
> **Barr Moses — CEO of Monte Carlo Data**

> "
> # We've got systems that work really well, but the ethical problems they create are burgeoning."
>
> **Jack Clark — Co-Founder of Anthropic, and co-chair of the AI Index Steering Committee**

As AI systems become more advanced and widespread, there is a growing concern about their ability to produce outputs that are out of line with human values. This is particularly evident in the case of generative AI, which has been known to generate untrue or harmful content. Naturally, the field of AI ethics has gained traction in covering these issues, with an increasing number of peer-reviewed papers being published on topics such as interpretability, explainability, causation, fairness, bias, and privacy.

In addition, the availability of tools that enable data teams to assess model fairness, explain model outputs, and mitigate bias has grown in popularity in the past few years. Examples of such tools include Google's Model Card Toolkit, FairML, and SHAP values.

These tools are especially useful when working on traditional supervised machine learning models, as they can enable organizations that use machine learning today to mitigate harmful outcomes from their models.

While the issues surrounding AI ethics remain complex and difficult to solve, we expect that awareness of these issues will continue to grow in 2023, as more companies and institutions take concrete steps to implement AI standards and ensure the responsible use of AI.

DIVE DEEPER INTO THE MODERN DATA STACK

Value Creation Within the Modern Data Stack — Learn key insights on how to the data tooling space is evolving from Chief Strategy Officer at Snowplow, Yali Sassoon.

Empowering the Modern Analyst — Learn how the modern data stack enables faster data insights from Peter Fishman, Co-Founder of Mozart Data.

Interpretable Machine Learning — Learn from Serg Masis, author of Interpretable Machine Learning with Python, how practitioners can build interpretable and explainable machine learning models.

The Future of Responsible AI — What does the future hold for responsible and interpretable machine learning? Maria Luciana Axente, Responsible AI and AI for Good Lead at PwC UK explains in this episode of DataFramed.

Innovation in tooling puts conceptual knowledge at the center

SKILLS & PEOPLE **SOCIETY & CULTURE**

In 2023, the rise of generative AI, and low to no-code tools will put the spotlight on the importance of data literacy and conceptual understanding of data workflows. While these tools make it easier for non-technical professionals to perform data analysis, build predictive models and communicate insights, they also require a deep understanding of data to use them effectively.

Data literacy, **the ability to read, work with, analyze, and communicate with data,** will become an increasingly valuable skill as the use of modern generative tools grows. Professionals will need to understand the basics of data concepts, such as data types, data structures, and data manipulation, to use these tools effectively.

Furthermore, a conceptual understanding of data workflows, from data collection and preparation to analysis and interpretation, will be essential for professionals to be able to design and implement effective data pipelines using generative AI and code generation tools.

This will require a deep understanding of the data analysis process, including data cleaning, feature engineering, model training, and evaluation.

In the longer term, the rise of low and no-code platforms could mean that proficiency in coding is no longer a strict prerequisite for data analytics. It will instead be data literacy that ranks high on the list of in-demand skills, along with communication skills and domain knowledge. In other words, everyone must gain expertise in exploring, understanding, and communicating with data. This speaks to the growing importance of digital upskilling in 2023.

8 Organizations and institutions continue to invest in data literacy

SKILLS & PEOPLE **SOCIETY & CULTURE** **DATA CULTURE**

The rise of data and machine learning has naturally sparked a need for workers with the skills needed to leverage these technologies effectively. In response, universities worldwide have begun offering data science programs to train the next generation of data professionals. These programs cover various topics, from statistics and computer science to data visualization and machine learning.

In addition to traditional academic institutions, many professional organizations are offering data literacy courses and certifications to help workers develop the skills they need to succeed in the rapidly changing world of data. For example, over 3,000 organizations are using DataCamp to upskill their workforce on data skills. Moreover, many organizations have built internal data academies that help accelerate their workforce's data skills.

In 2023 however, we'll see the inclusion of data literacy in school curricula. Countries like Estonia, Korea, and Kazakhstan have begun emphasizing the importance of digital and data literacy in their education systems. In the United States, the Department of Education has recommended teaching data literacy basics, such as understanding what data is and how it can be used to answer questions, from grade 1 onwards.

Overall, the increasing focus on data literacy is a positive development that will help workers stay competitive in the job market and enable organizations to make better, more data-driven decisions. We expect to see continued growth in this area in the coming years.

DIVE DEEPER INTO DATA LITERACY

How CBRE is Increasing Data Literacy for Over 3,000 Employees — Learn from Emily Hayward, Data & Digital Change Manager at CBRE, about how CBRE is driving data literacy for more than 3,000 people.

How Data Literacy Skills Help You Succeed — Jordan Morrow, Author of be data literate: The data literacy skills everyone needs to succeed, joins the show to discuss how organizations can drive data literacy, and how data skills drive career outcomes.

Data Literacy: A how-to Guide for Leaders — Join this webinar with Cindi Howson, Chief Data Strategy Officer at ThoughtSpot, to learn how you can start building data literacy within your organization.

Allianz Case Study — Read this case study to learn how Allianz upskilled more than 6,000 people within their internal data academy using DataCamp.

9 Recession-proof industries will be a safe haven for analytics professionals

The impact of the COVID-19 pandemic on the tech industry has been widespread, with many tech companies cutting back on the workforce investments they made throughout the pandemic. However, despite the challenges that the pandemic has brought, we remain optimistic about the long-term growth of the data industry.

The current economic downturn has caused a surplus of tech workers and negative impacts on salaries and job opportunities. However, the long-term demand for data talent will remain strong as companies increasingly rely on data-driven decision-making. While the tooling innovations we talked about throughout this white paper will indeed democratize data workflows for many today, the demand for these data skills will continue to grow.

Furthermore, we expect that the demand for data talent will be particularly strong in industries that are resistant to economic downturns. Sectors such as healthcare, transportation, and necessity retailers are likely to continue to need data professionals to help them make strategic decisions and optimize their operations.
In conclusion, despite the challenges brought about by the current economic downturn, the demand for data talent will remain strong in the long term as companies continue to rely on data-driven decision-making. The rise of new technologies, such as AI and machine learning, will also create new opportunities for data professionals to apply their skills in innovative ways.

10 The rise of specialist new roles in data signals a maturing role ecosystem

SKILLS & PEOPLE   **JOBS & HIRING**

While data roles such as data analyst and data scientist are maturing and becoming more clearly defined, the broader data job landscape is still in flux, with job definitions changing and new roles appearing according to business needs.

One of the most notable trends in recent years has been the rise of the analytics engineer, a role that sits between a data engineer and a data analyst. Analytics engineers possess a unique combination of domain knowledge and data engineering experience, allowing them to modify and optimize existing pipelines to address the common knowledge gap between data engineers and data analysts.

In addition to these specialized roles, there has also been a rise in demand for data communicators, data project managers, and data product managers. As data literacy becomes increasingly important in organizations across all industries, we predict that these roles will continue to gain adoption in 2023.

DIVE DEEPER INTO THE DATA JOB MARKET

How Data Leaders Can Build an Effective Talent Strategy — Kyle Winterbottom, founder of Orbition and Data IQ's 100 Most Influential People in Data for 2022, talks about the key differentiators between data teams that build talent-dense teams and those that do not.

Hiring & Retaining Data Talent — Learn from Meenal Iyer, VP of Data at Momentive.ai, and Glenn Hofmann, Chief Analytics Officer at New York Life Insurance, about key principles for hiring and retaining data talent in the years to come.

The Rise of Hybrid Jobs & The Future of Data Careers — Matt Sigelman, Chairman of Lightcast and president of The Burning Glass Institute describes how data roles are evolving, and data skills are leading to hybrid jobs.

Employer Branding for Data Talent When You're Not a FAANG — Learn from data recruiting leaders how to optimize a data talent strategy when you're not a major technology firm

About DataCamp

DataCamp is a one-stop shop for building a data-driven workforce. With DataCamp for Business, you can transform how everyone in your organization uses data. With more than 380+ interactive, self-paced online courses, a cloud-based IDE that connects to your company data, and a recruitment platform for hiring data talent, DataCamp for Business will help you make the most of your organization's data.