From Data Strategy to Knowledge Strategy: The Future of the Artificial Intelligence Revolution

Eviatar Matania and Tomer Y. Avni

We would like to thank David J. Blumberg, Jacob Rimer and Ori Rozen for contributing to this article

The Artificial Intelligence (AI) revolution, ignited more than half a century ago, is based on data and driven by knowledge. In the last decade, AI has grown from an academic scientific field to start being a practical part of our everyday lives. The most common AI business strategies we see are built around data. Yet, due to advancements in the ability and willingness to exchange data, we believe that within a decade proprietary data moats will be less sustainable. Therefore, we expect a shift in focus, from data-based AI strategies, to knowledge-based AI strategies. In this article, we offer key ways to prosper amid this strategic shift.

The Artificial Intelligence (AI) revolution was ignited more than half a century ago. In the last decade, AI has grown from an academic scientific field to begin being a practical part of our everyday lives and it is being used in production by many businesses.[1] The most common AI business strategies we see are built around data. We believe that proprietary data is currently the most strategic moat for AI companies, but in the coming years, it will become less of a unique asset, making proprietary data differentiation less sustainable. Therefore, we expect a shift in focus, from data-based AI strategies, to knowledge-based AI strategies. It follows that a strategy overhaul is required for businesses to prosper as AI continues to develop and increasingly impact our lives.

While scientific research in the field of AI started more than half a century ago, it was only in the last 10-15 years that AI began having a noticeable and dramatic impact in business and other domains. The confluence of three major technological breakthroughs enabled this disruption. First,

[1] "Big Data and AI Executive Survey 2019," January 2019, http://newvantage.com/wp-content/uploads/2018/12/Big-Data-Executive-Survey-2019-Findings-Updated-010219-1.pdf

the rapid rise in the deployment and use of low-cost sensors, (e.g. endpoint sensors in cellphones, watches, cameras and smart home devices), together with various industrial sensors that generate massive data streams (i.e. big data). Second, the ubiquity of Internet connectivity. Third, the development of cost-effective software and hardware systems for computational power, (i.e. Graphical Processor Units (GPUs)), with communication abilities that enabled efficient data transmission and digital storage. These combined breakthroughs of data, connectivity and computational hardware and software have enabled AI to scale from small academic research projects to large enterprise production applications. Essentially, big data required sophisticated AI models to analyze and derive knowledge and insights, while the AI models needed the critical mass of "big data" for training and optimization.[2]

At present, many startups place data acquisition at the heart of their business strategy and perceive data as a strategic moat. As venture capital investors, we have seen an increasing number of companies emphasizing the unique data sets they have acquired and their long-term strategy for acquiring additional proprietary data - as a sustainable barrier of entry. Moreover, As AI tools and AI-as-a-service platforms have commoditized the development of AI models and public data has become ubiquitous, the perceived need to build and defend a data moat has become palpable.

Indeed, in today's technology ecosystem, the markets have increasingly rewarded companies with leading AI programs and control over proprietary data - as a substantial and sustainable competitive advantage. Companies such as Google and Netflix have developed and curated massive and authoritative datasets over a long period of time, while many other companies struggled in vain to match their success. An example is the massive disruption of rival media service providers and production companies, which were outmaneuvered by Netflix' sophisticated data strategy.[3]

---

[2] Randy Bean, "How Big Data Is Empowering AI and Machine Learning at Scale," May 8, 2017,https://sloanreview.mit.edu/article/how-big-data-is-empowering-ai-and-machine-learning-at-scale

[3] Ritwik Kumar, Vinith Misra, Jen Walraven, Lavanya Sharan, Bahareh Azarnoush, Boris Chen, Nirmal Govind, "Data Science and the Art of Producing Entertainment at Netflix," March 26, 2018, https://netflixtechblog.com/studio-production-data-science-646ee2cc21a1?gi=4f5246c690cb ; Netflix Research, https://research.netflix.com/research-area/analytics ; Enrique Dans, "Netflix: Big Data And Playing A Long Game Is Proving A Winning Strategy," January 15, 2020, https://www.forbes.com/sites/enriquedans/2020/01/15/netflix-big-data-and-playing-a-long-game-is-proving-a-winningstrategy#5ac5ac84766e ; Jon Markman, "Netflix Harnesses Big Data To Profit From Your Tastes,"

Nevertheless, due to expected advancements in the ability and willingness to exchange data, we believe that within a decade proprietary data moats will be less sustainable. While data will still fuel the AI value engine, AI business strategies will be increasingly focused on knowledge.

Moving Up the AI Value Pyramid, towards the Knowledge Layer

The AI value pyramid is based on data and driven by knowledge. Data is the essential raw material obtained through a wide variety of sources. Then, various cleansing and structuring techniques and algorithms are employed to transform this raw data into structured data or information that is ready for use in AI models. Finally, the successful AI models use the information to generate valuable knowledge and insights.[4] We observe this flow in numerous verticals. For example, in the mobility space, the Israeli navigation company Waze (acquired by Google), collects raw data from drivers around the world and other external sources, transforms it to information, by structuring and labeling this data and then applies its models on this information, to generate insights for drivers in real time. Similar examples can be found in other domains, including the Fintech and Cybersecurity domains.

This AI value pyramid underlies our vision of the pace and direction of AI's continuing evolution and how it will develop in coming decades. The structure of the AI value pyramid resembles significant periods of advancement in human history, including the industrial revolution. The industrial revolution commenced when engineering innovations, such as the steam engine, were able to transform lumps of carboniferous black rock (coal) fuel into energy that powered the production of goods and services at an unprecedented scale, speed and at a fraction of the cost of traditional manual methods. Although the industrial revolution's value was initially generated through the utilization of coal as fuel, over time, greater value shifted to the product and services

February 25, 2019, https://www.forbes.com/sites/jonmarkman/2019/02/25/netflix-harnesses-big-data-to-profit-from-your-tastes ; and Enrique Dans, "Why Everybody Wants To Be Like Netflix," July 11, 2018, https://www.forbes.com/sites/enriquedans/2018/07/11/why-everybody-wants-to-be-like-netflix#399e07db7fd8

[4] A. Aamodt and M. Nygard, "Different roles and mutual dependencies of data, information, and knowledge — An AI perspective on their integration," Data & Knowledge Engineering 16 (September 1995): 191-222.

layer while the raw fuel and even the harnessed energy became commodities. Looking back at the industrial revolution from a contemporary point of view exposes the product and services layer as its value linchpin. This is clearly demonstrated by analyzing the GDP composition of developed countries. For example, energy and fuel production constitute less than 20 percent of the US GDP, while product and services contribute more than 80 percent.[5] Over time, fuel and energy became commoditized, yet even when the OPEC organization attempted to make petroleum fuel more scarce and "proprietary" with production and sale quotas, new innovative methods of production ("fracking" technology) were invented and returned fuel to its commodity status.

Today, fuel ownership is not as significant as product development. We believe the same will happen with the future evolution of AI. While fuel can be burned and used once, the same data set can be used over and over by many customers. There is a proliferation of data sources, some overlap and yield two sources for equivalent dataset, which reduces uniqueness and the proprietary barrier. In addition, there are new techniques, protocols and standards for pooling, sharing and exchanging data that make it more of a commodity, and less proprietary over time.

While today "we are drowning in information but starved for knowledge",[6] we expect moving up the AI value chain, towards the knowledge layer. Indeed, we have begun to see advances that will foster and accelerate this trend by creation of data exchanges. We expect that data exchange will be facilitated by a combination of increased feasibility and willingness to share commoditized data in return for valuable knowledge. In summary, data will become more plentiful, available, reliable and standardized and inexpensive - the perfect definition of an ideal commodity. Therefore, using data as a sustainable barrier of entry will be more difficult in the future.

The growth of standardized data and the proliferation of data sharing methodologies and tools will enhance the ability to share proprietary data. The Internet of Things (IoT) domain[7] is rapidly

[5] Central Intelligence Agency, "The World Factbook", https://www.cia.gov/library/publications/the-world-factbook/fields/214.html

[6] John Naisbitt and Particia Aburdene, "Megatrends 2000 New Directions for Tomorrow" (New York: Warner Books, 1984).

[7] B. Muhammad, R. Rana Asif, and K. Byung-Seo, "IoT Elements, Layered Architectures and Security Issues: A Comprehensive Survey," Sensors 18 (August 2018); "Internet of Things (IoT) connected devices installed base

generating vast amounts of data collected in almost every vertical from manufacturing, to medical care to intelligent infrastructure, and autonomous vehicles.

We have already seen the rise of data standardization[8] in recent years. An example of this is the European Committee for Standardization's data framework, known as Transmodel, for public transportation.[9] Another example is Project Connected Home over IP.[10] This development by a consortium of leading technology companies aims to create data and connectivity standards across smart home sensors and other devices for the common benefit of IoT vendors around the world. Furthermore, governments are also intervening to compel companies and organizations to make data public and available in standard formats. For example, the PSDII standards in the EU require banks to expose and share their financial data with a myriad of data aggregators and vendors, who use that data to compete with the very same banks. The more standardized the data will be, the easier it will be to share data across organizations and the more it will be commoditized, and then turned into valuable knowledge using the catalyst of AI.

The surge in the volume of standardized data will be complemented by the invention of new ways to share data. Inventions such as Google's Federated Learning,[11] Homomorphic Encryption[12] for

---

worldwide from 2015 to 2025," https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide ; and Laurence Goasduff, "Gartner says 5.8 Billion Enterprise and Automotive IoT Endpoints Will Be in Use in 2020," August 29, 2019, https://www.gartner.com/en/newsroom/press-releases/2019-08-29-gartner-says-5-8-billion-enterprise-and-automotive-io

[8] Frederic Lardinois, "Microsoft, SAP and Adobe take on Salesforce with their new Open Data Initiative for customer data," September 24, 2018, https://techcrunch.com/2018/09/24/microsoft-sap-and-adobe-take-on-salesforce-with-their-new-open-data-initiative-for-customer-data ; Frederic Lardinois, "Microsoft, Adobe and SAP prepare to expand their Open Data Initiative," March 27, 2019, https://techcrunch.com/2019/03/27/microsoft-adobe-and-sap-prepare-to-expand-their-open-data-initiative ; Mary Jo Foley, "What's next for the Open Data Initiative?," November 18, 2019, https://www.zdnet.com/article/whats-next-for-the-open-data-initiative ; and The Open Data Initiative, https://www.microsoft.com/en-us/open-data-initiative

[9] Transmodel, http://www.transmodel-cen.eu/

[10] Project Connected Home over IP, https://www.connectedhomeip.com/

[11] Brendan McMahan and Daniel Ramage, "Federated Learning: Collaborative Machine Learning without Centralized Training Data," April 6, 2017, https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

[12] Lukas Rist, "Encrypt your Machine Learning,", January 8, 2018, https://medium.com/corti-ai/encrypt-your-machine-learning-12b113c879d6 ; and Keno Fischer, "Machine Learning on Encrypted Data Without Decrypting It," November 22, 2019, https://juliacomputing.com/blog/2019/11/22/encrypted-machine-learning.html

AI, Differential Privacy[13] and others, will enable multiple parties to share proprietary data and utilize it to train their models without exposing the data to others. Furthermore, tools for data sharing, including AWS data exchange[14] among others[15] will also make proprietary data sharing more feasible.

The increased ability to share data will become truly significant when there is incentive and a growing inclination to do so. As AI undermines and disrupts legacy competitive barriers to entry,[16] many organizations relentlessly attempt to collect their own proprietary data and monetize it. Alas, this data acquisition and utilization is increasingly neither easy, nor fruitful and therefore creates strategic dissonance. This is because while AI is increasingly indispensable for most organizations, it's not part of their legacy skills or core expertise. In addition, the chronic and enduring shortage[17] of engineers, developers, product leads and managers trained in AI, sharpens this dissonance and leads to a solution preference for data sharing with the goal of knowledge exchange.

An example of the combination of ability and willingness creating the exchange of data for knowledge generation is the new proposal[18] by the European Union, to create "a single market for

---

[13] Georgian Partners, "A Brief Introduction to Differential Privacy," August 31, 2018, https://medium.com/georgian-impact-blog/a-brief-introduction-to-differential-privacy-eacf8722283b ; Ria Cheruvu, "A High Level Introduction to Differential Privacy," November 19, 2018, https://towardsdatascience.com/a-high-level-introduction-to-differential-privacy-edd20e6adc3b ; and An Nguyen, "Understanding Differential Privacy," June 30, 2019, https://towardsdatascience.com/understanding-differential-privacy-85ce191e198a

[14] AWS Data Exchange, https://aws.amazon.com/data-exchange/ ; and Kyle Wiggers, "Amazon launches AWS Data Exchange for tracking and sharing data sets,", November 13, 2019, https://venturebeat.com/2019/11/13/amazons-aws-data-exchange-launches-with-over-80-data-providers/

[15] Midata, https://www.midata.coop/en/home/ ; Zack Whittaker, "A group of ex-NSA and Amazon engineers are building a 'Github for data'," February 20, 2020, https://techcrunch.com/2020/02/20/gretel-nsa-amazon-github-data/ ; V. Molina, M. Kersten-Oertel, and T. Glatard, "A Conceptual Marketplace Model for IoT Generated Personal Data," July 5, 2019, https://arxiv.org/pdf/1907.03047.pdf ; and M. Benjamin, P. Gagnon, N. Rostamzadeh, C. Pal, Y. Bengio, and A. Shee, "Toward Standardization of Data Licenses: the Montreal Data License," March 21, 2019, https://arxiv.org/pdf/1903.12262.pdf

[16] M. Iansiti and K. R. Lakhani, "Competing in the Age of AI," Harvard Business Review 98 (January-February 2020): 60-67; and "Big Data and AI Executive Survey 2019," January 2019, http://newvantage.com/wp-content/uploads/2018/12/Big-Data-Executive-Survey-2019-Findings-Updated-010219-1.pdf

[17] Cade Metz, "Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent," October 22, 2017, https://www.nytimes.com/2017/10/22/technology/artificial-intelligence-experts-salaries.html

[18] European Data Strategy, https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy

data" so that "people, business and organizations… be empowered to make better decisions based in insights from non-personal data", in order to compete with the current tech giants.[19]

Finally, another factor contributing to data moats becoming less sustainable is the invention of novel data solutions which enable using smaller sets of data for training models. Synthetic data solutions (for example, with Generative Adversarial Networks) and other minimization techniques, like data augmentation, might allow companies to create disruptive AI products, without huge amounts of data.

Building a Knowledge Strategy

The future of the AI revolution will usher in a new reality for businesses and will require revised business strategy. The shift from data to knowledge will generate novel frameworks, partnerships and business models, encompassing different players bringing data, information, AI models, storage, and computing power, for knowledge creation. Since traditional data moats will become less sustainable in the coming decade, and knowledge will become the real value driver of AI, we believe businesses should begin crafting a strategy that is more focused on knowledge:

- Building knowledge moats, rather than data moats, is a fundamental principle that should be at the heart of future business strategy. Companies and organizations should start preparing for a knowledge-centric era, in which the winners will be those asking the right questions, looking for the most relevant predictions and designing the most disruptive AI-based applications.

- Use AI in a top-down manner and structure businesses around the application and product layer. Models should be developed and trained based on the specific vertical and hypothesis. An example would be to develop specific health care applications in imaging, diagnostics, telemedicine, pharmacology and other clinical applications; or in mobility across fleet management, public transportations, and beyond. The development of these

[19] Foo Yun Chee, "Europe wants single data market to break U.S. tech giants' dominance," January 29, 2020, https://www.reuters.com/article/us-eu-data-exclusive/exclusive-europe-wants-single-data-market-to-break-u-s-tech-giants-dominance-idUSKBN1ZS32E

solutions will be based on deep knowledge and practical experience in specific domains, combining contextual knowledge and appropriate and well-tuned models.

- Data acquisition initiatives should be viewed as merely a short-term tactical pursuit while knowledge-based partnerships for exchange and cooperation should be fostered and cultivated as long-term business strategies. A productive example is that last year, the Israeli Innovation Authority launched a pilot program for knowledge-based cooperation between hospitals and technology startups. This cooperation generated dozens of initiatives[20] between hospitals of startups and facilitated the exchange of raw (and mostly unused) data from and among the hospitals, and generation of novel and valuable knowledge generated by the startups.

- Lastly, the shift towards knowledge should shape organizations' HR strategy as well. Companies should develop a relevant and smart HR strategy for the future of AI. While some startups will still require hiring large cohorts of rare and expensive data engineers and scientists, savvy companies' AI teams should instead be designed as a managerial group designed to pursue and foster AI knowledge partnerships, invent AI-based applications and products and creatively explore the bright horizons of the AI revolution - reimagined from data-centric to knowledge-centric. Furthermore, AI teams should have people who understand the context of the domain they are operating in. These contextual team members should encompass a holistic approach, originated in their understanding of AI and the specific domain, and not only general AI experts.

Overall, the future of AI depends on shifting from emphasis on proprietary data sets, to the sharing of data across entities for knowledge creation. To implement a successful AI strategy, companies must rightly combine data, information, AI models, storage, computing power and more, in order to root their business in knowledge.

A shorter version of this article was published in September 2020 on The Next Web

[20] Deputy Head of the Israeli Innovation Authority, "Data Regime for Israel" Conference at Tel Aviv University, February 2020.

About the Authors

Eviatar Matania is a professor at the School of Political Sciences, Government and International Affairs at Tel-Aviv University, where he heads the MA program of Security Studies and the MA program of cyber politics and government (under inception). Matania is also an Adjunct Professor at Oxford's Blavatnik School of Government. Matania was the founding head and former Director General of the Israel National Cyber Directorate (INCD) reporting directly to the Israeli Prime Minister, and During the last two years he has been co-heading the Smart Systems (AI) National Initiative to recommend the Prime Minister and the government about a national plan to posture Israel as a global power in AI.

Tomer Y. Avni is an MS/MBA candidate at Harvard Business School and the Harvard School of Engineering and Applied Sciences.

Bibliography

"Big Data and AI Executive Survey 2019," January 2019, http://newvantage.com/wp-content/uploads/2018/12/Big-Data-Executive-Survey-2019-Findings-Updated-010219-1.pdf

Randy Bean, "How Big Data Is Empowering AI and Machine Learning at Scale," May 8, 2017,https://sloanreview.mit.edu/article/how-big-data-is-empowering-ai-and-machine-learning-at-scale

Ritwik Kumar, Vinith Misra, Jen Walraven, Lavanya Sharan, Bahareh Azarnoush, Boris Chen, Nirmal Govind, "Data Science and the Art of Producing Entertainment at Netflix," March 26, 2018, https://netflixtechblog.com/studio-production-data-science-646ee2cc21a1?gi=4f5246c690cb

Netflix Research, https://research.netflix.com/research-area/analytics

Enrique Dans, "Netflix: Big Data And Playing A Long Game Is Proving A Winning Strategy," January 15, 2020, https://www.forbes.com/sites/enriquedans/2020/01/15/netflix-big-data-and-playing-a-long-game-is-proving-a-winningstrategy#5ac5ac84766e

Jon Markman, "Netflix Harnesses Big Data To Profit From Your Tastes," February 25, 2019, https://www.forbes.com/sites/jonmarkman/2019/02/25/netflix-harnesses-big-data-to-profit-from-your-tastes

Enrique Dans, "Why Everybody Wants To Be Like Netflix," July 11, 2018, https://www.forbes.com/sites/enriquedans/2018/07/11/why-everybody-wants-to-be-like-netflix#399e07db7fd8

A. Aamodt and M. Nygard, "Different roles and mutual dependencies of data, information, and knowledge — An AI perspective on their integration," Data & Knowledge Engineering 16 (September 1995): 191-222.

Central Intelligence Agency, "The World Factbook", https://www.cia.gov/library/publications/the-world-factbook/fields/214.html

John Naisbitt and Particia Aburdene, "Megatrends 2000 New Directions for Tomorrow" (New York: Warner Books, 1984).

B. Muhammad, R. Rana Asif, and K. Byung-Seo, "IoT Elements, Layered Architectures and Security Issues: A Comprehensive Survey," Sensors 18 (August 2018); "Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025," https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide

Laurence Goasduff, "Gartner says 5.8 Billion Enterprise and Automotive IoT Endpoints Will Be in Use in 2020," August 29, 2019, https://www.gartner.com/en/newsroom/press-releases/2019-08-29-gartner-says-5-8-billion-enterprise-and-automotive-io

Frederic Lardinois, "Microsoft, SAP and Adobe take on Salesforce with their new Open Data Initiative for customer data," September 24, 2018, https://techcrunch.com/2018/09/24/microsoft-sap-and-adobe-take-on-salesforce-with-their-new-open-data-initiative-for-customer-data

Frederic Lardinois, "Microsoft, Adobe and SAP prepare to expand their Open Data Initiative," March 27, 2019, https://techcrunch.com/2019/03/27/microsoft-adobe-and-sap-prepare-to-expand-their-open-data-initiative

Mary Jo Foley, "What's next for the Open Data Initiative?," November 18, 2019, https://www.zdnet.com/article/whats-next-for-the-open-data-initiative
The Open Data Initiative, https://www.microsoft.com/en-us/open-data-initiative

Transmodel, http://www.transmodel-cen.eu/

Project Connected Home over IP, https://www.connectedhomeip.com/

Brendan McMahan and Daniel Ramage, "Federated Learning: Collaborative Machine Learning without Centralized Training Data," April 6, 2017, https://ai.googleblog.com/2017/04/federated-learning-collaborative.html

Lukas Rist, "Encrypt your Machine Learning,", January 8, 2018, https://medium.com/corti-ai/encrypt-your-machine-learning-12b113c879d6

Keno Fischer, "Machine Learning on Encrypted Data Without Decrypting It," November 22, 2019, https://juliacomputing.com/blog/2019/11/22/encrypted-machine-learning.html

Georgian Partners, "A Brief Introduction to Differential Privacy," August 31, 2018, https://medium.com/georgian-impact-blog/a-brief-introduction-to-differential-privacy-eacf8722283b

Ria Cheruvu, "A High Level Introduction to Differential Privacy," November 19, 2018, https://towardsdatascience.com/a-high-level-introduction-to-differential-privacy-edd20e6adc3b

An Nguyen, "Understanding Differential Privacy," June 30, 2019, https://towardsdatascience.com/understanding-differential-privacy-85ce191e198a

AWS Data Exchange, https://aws.amazon.com/data-exchange/

Kyle Wiggers, "Amazon launches AWS Data Exchange for tracking and sharing data sets,", November 13, 2019, https://venturebeat.com/2019/11/13/amazons-aws-data-exchange-launches-with-over-80-data-providers/

Midata, https://www.midata.coop/en/home/

Zack Whittaker, "A group of ex-NSA and Amazon engineers are building a 'Github for data'," February 20, 2020, https://techcrunch.com/2020/02/20/gretel-nsa-amazon-github-data/

V. Molina, M. Kersten-Oertel, and T. Glatard, "A Conceptual Marketplace Model for IoT Generated Personal Data," July 5, 2019, https://arxiv.org/pdf/1907.03047.pdf

M. Benjamin, P. Gagnon, N. Rostamzadeh, C. Pal, Y. Bengio, and A. Shee, "Toward Standardization of Data Licenses: the Montreal Data License," March 21, 2019, https://arxiv.org/pdf/1903.12262.pdf

M. Iansiti and K. R. Lakhani, "Competing in the Age of AI," Harvard Business Review 98 (January-February 2020): 60-67

"Big Data and AI Executive Survey 2019," January 2019, http://newvantage.com/wp-content/uploads/2018/12/Big-Data-Executive-Survey-2019-Findings-Updated-010219-1.pdf

Cade Metz, "Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent," October 22, 2017, https://www.nytimes.com/2017/10/22/technology/artificial-intelligence-experts-salaries.html

European Data Strategy, https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy

Foo Yun Chee, "Europe wants single data market to break U.S. tech giants' dominance," January 29, 2020, https://www.reuters.com/article/us-eu-data-exclusive/exclusive-europe-wants-single-data-market-to-break-u-s-tech-giants-dominance-idUSKBN1ZS32E

Deputy Head of the Israeli Innovation Authority, "Data Regime for Israel" Conference at Tel Aviv University, February 2020.