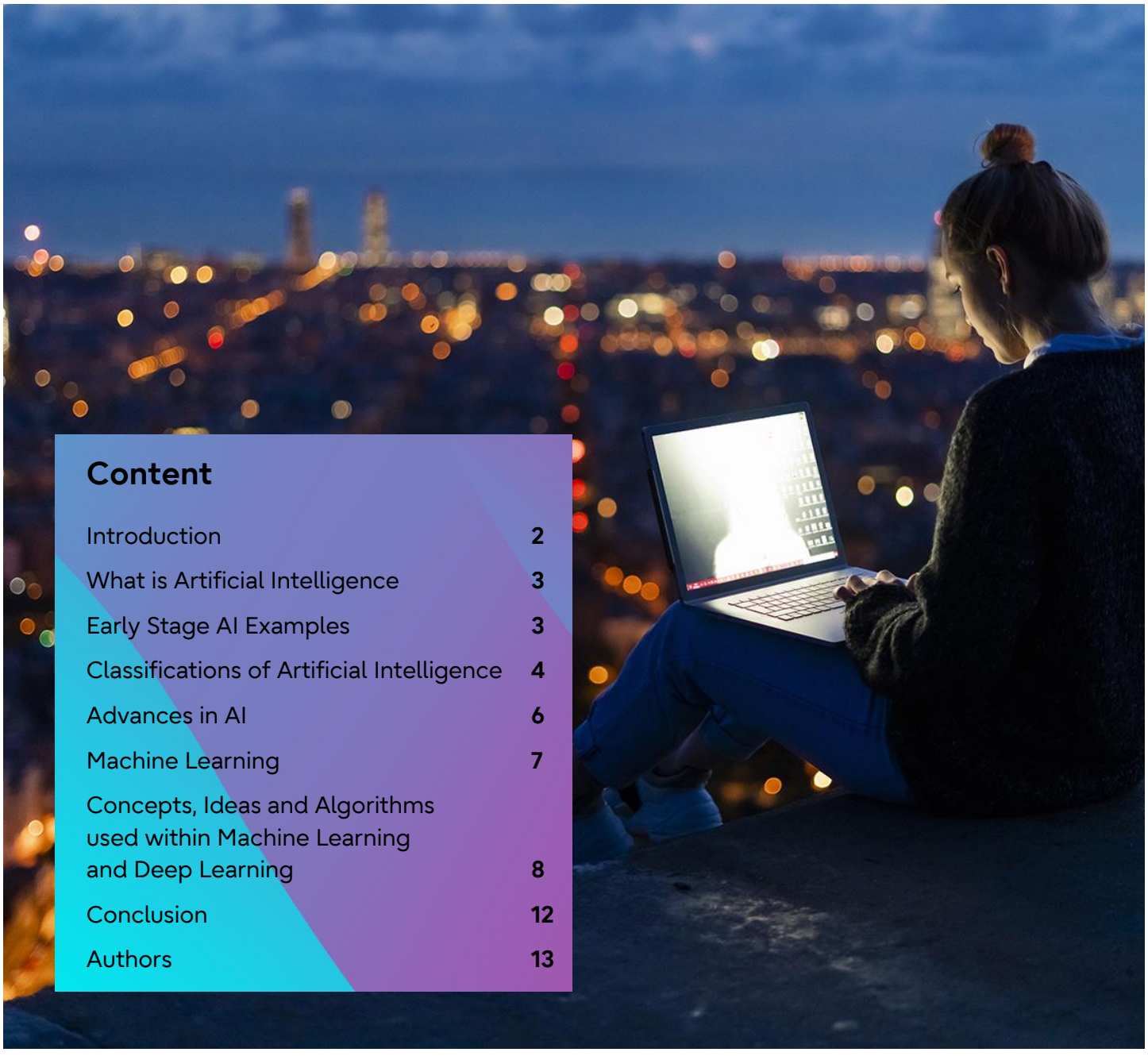
FUJITSU

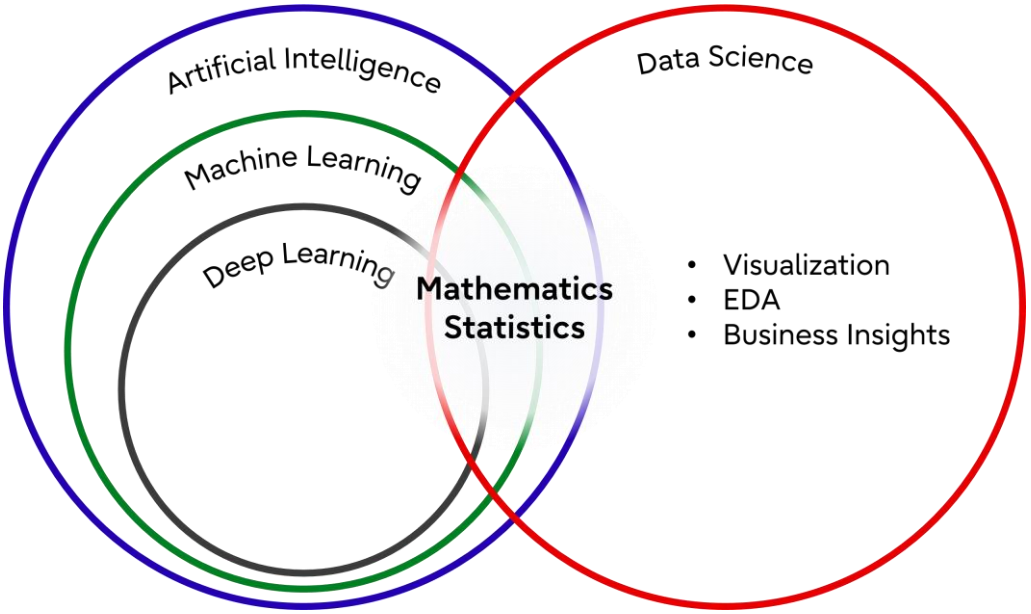
AI and Advanced Analytics
A plain speaking guide

A foundational guide on technical aspects of AI and Advanced Analytics, and how they are driving change in business and in our day-to-day lives



Content

Introduction



It is highly likely humanity will need Artificial Intelligence (AI) to help address existential needs, such as sustainable solutions to the climate emergency. But without trust in AI - and technology more generally – these solutions are likely to be delayed or even abandoned. Many people feel threatened by AI – at a time in human society when we have a pressing need for the advances it can offer.

To help build trust in this important technology, the authors of this paper decided to create a "grounded and rounded" view of where AI and advanced analytics are today. One that does not talk down to the novice and does not over-promise, over-claim, or paper over the areas where AI is currently weak. We have aimed to cover all the essential aspects of AI, leaving you to pick your passion and drill down into the information you need. We cover the technical aspects of AI and advanced analytics, but we also put this into context in terms of how they are driving change in business and our day-to-day lives. The information will enable you to transition from novice – if that is where you are today – to enthusiast.

What is Artificial Intelligence?

Intelligence is the ability to learn, understand, and make judgments or have opinions that are based on reason.

Artificial is something that is made by people, often as a copy of something natural.

AI therefore refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind, such as learning and problem-solving.

The ideal characteristic of AI is its ability to rationalize and take actions that have the best chance of achieving a specific goal.

Early-Stage AI Examples

Example 1
In 2011, conservationists were interested in the noise produced when a bird collides with a power line. With some 600 hours of audio collected in power line stations (25 days' worth of audio clips), manually counting the bird collisions was impractical. So, the audio files (as well as metadata such as times and locations), were sent to a company that uses AI to assist wildlife monitoring. They were able to detect the collisions automatically. Over several years, results suggested that bird deaths numbered in the high hundreds or low thousands, much higher than expected, and providing far more accurate quantification of the true impact of power lines on endangered species.

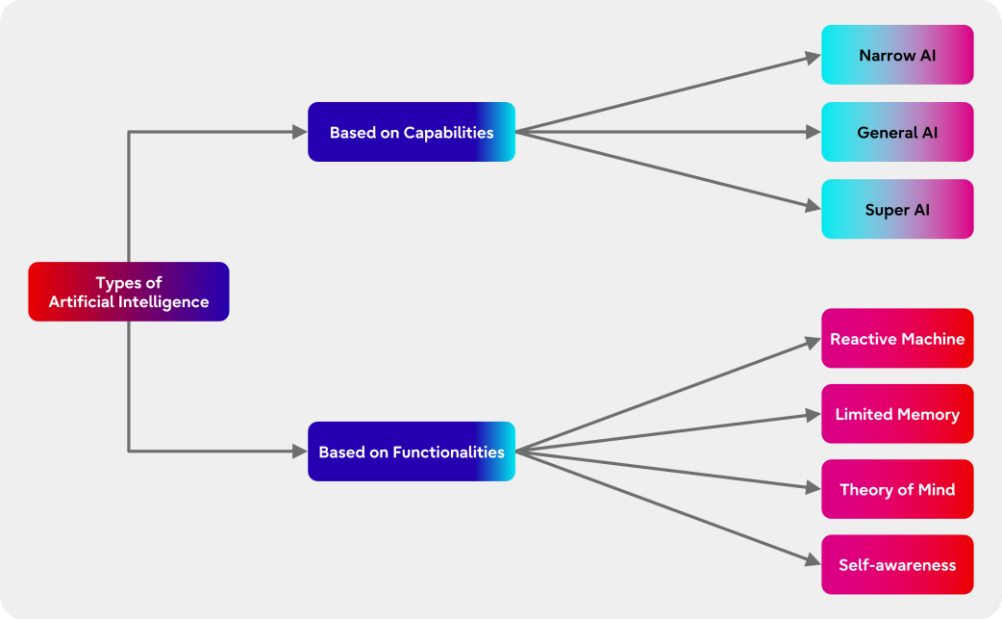
Example 2
Law enforcement agencies pursuing illegal wildlife traffickers need to determine quickly if an animal for sale on social media is protected or not. AI provides a rapid alternative to manually processing huge troves of data, such as camera-trap images or audio recordings. A PhD student could spend months hand-labelling data before reaching valid conclusions.

Example 3
To enable more precise estimates of wildlife population sizes, neural networks and computer-vision algorithms can be used to detect animals in images, count them and identify individual types of animals within a species.

Example 4
In the financial services sector, software is taking decisions about loans based on a variety of finely parsed data about a borrower, rather than just a credit score and a background check. Robot advisers can also create personalized investment portfolios, avoiding the need for stockbrokers and financial advisers. Advances in AI also allow stock traders to take the emotion out of investing and take rapid decisions based on analytical considerations.

Example 5
AI is helping financial services organizations overcome the difficulty of detecting fraudulent activities in large companies. AI can identify abnormalities, outliers or deviant cases that require additional investigation, helping managers find problems early in the cycle before they reach dangerous levels.

Classifications of Artificial Intelligence



There are many classifications of AI systems, of which two are prominent: one based on capabilities and the other based on functionalities.

Classification based on capabilities

The classification in general used in tech parlance is based on capabilities:

1. **Artificial Narrow Intelligence (ANI) – Task-Based**

This type represents all existing AIs. Even the most complex AIs ever created, including those using machine learning and deep learning to teach themselves, fall under ANI. Artificial Narrow Intelligence refers to all reactive and limited-memory AI systems that can only perform a specific task autonomously using human-like capabilities. These machines can do nothing more than what they are programmed to do, and thus have a very limited or narrow range of competencies.

2. **Artificial General Intelligence (AGI) - Learns by Itself**

Artificial General Intelligence is the ability of an AI agent to learn, perceive, understand, and function like a human being. It does not yet exist. These systems will be able to independently build multiple competencies and form connections and generalizations across domains, massively cutting down on time needed for training. This will make AI systems just as capable as humans by replicating our multi-functional capabilities.

3. **Artificial Superintelligence (ASI)**

The development of Artificial Superintelligence will probably mark the pinnacle of AI research, as ASI will become by far the most capable form of intelligence on Earth. ASI, in addition to replicating the multi-faceted intelligence of human beings, will be far better at everything we do because of overwhelmingly greater memory, faster data processing and analysis, and decision-making capabilities. The development of AGI and ASI will lead to a scenario most popularly referred to as the singularity. And while the potential of having such powerful machines at our disposal seems appealing, these machines may also threaten our existence or at the very least, our way of life.

Classification based on functionalities

The classification in general used in tech parlance is based on functionalities:

4. **Reactive Machines**

Reactive Machines are the oldest forms of AI and have extremely limited capabilities. They emulate the human mind's ability to respond to different kinds of stimuli. These machines do not have memory-based functionality. This means they cannot use previous experiences to inform their present actions, i.e., these machines do not have the ability to "learn". Uses are restricted to automatically responding to a limited set or combination of inputs and they cannot rely on memory to improve their operations. A popular example of a reactive AI machine is IBM's Deep Blue, a machine that beat chess Grandmaster Garry Kasparov in 1997.

5. **Limited Memory Machines**

In addition to the capabilities of purely reactive machines, Limited Memory Machines are also capable of learning from historical data to make decisions. Nearly all existing AI applications that we know of - from chatbots and virtual assistants to self-driving vehicles - come under this category. All present-day AI systems, such as those using deep learning, are trained by large volumes of data that they store in their memory to form a reference model for solving future problems. For instance, an image recognition AI is trained using thousands of pictures and their labels to teach it to name other objects. When a new image is scanned by this type of AI, it uses the training images as references to understand the content and, based on its prior "learning experience", it labels new images with increasing accuracy.

6. **Theory of Mind**

While the previous two types of AI are abundant, the next two types exist, for now, either as a concept or a work in progress. Theory of mind AI is the next level of AI that researchers are currently innovating. A theory of mind level AI will be able to better understand the entities it is interacting with by discerning their needs, emotions, beliefs, and thought processes. While artificial emotional intelligence is already a budding industry and an area of interest for leading AI researchers, achieving theory of mind AI will require development in other branches of AI as well. This is because, to truly understand human needs, AI machines will have to perceive humans as individuals whose minds can be shaped by multiple factors, essentially "understanding" us.

7. **Self-Aware Systems**

This is the final stage of AI development, which currently exists only hypothetically. Self-explanatory, it is an AI that has evolved to be so akin to the human brain that it has developed self-awareness. Creating this type of AI, which is decades, if not centuries away from materializing, is and will always be the ultimate objective of all AI research. This type of AI will not only be able to understand and evoke emotions in those it interacts with, but also have emotions, needs, beliefs, and potentially desires of its own. And this is the type of AI that technology doomsayers are wary of. The development of self-aware AI could potentially boost our progress as a civilization by leaps and bounds. Once self-aware, the AI would be capable of having human-like conversations, becoming the ideal companion in our day-to-day tasks, responding to our emotional needs in deep private conversations.

Advances in AI

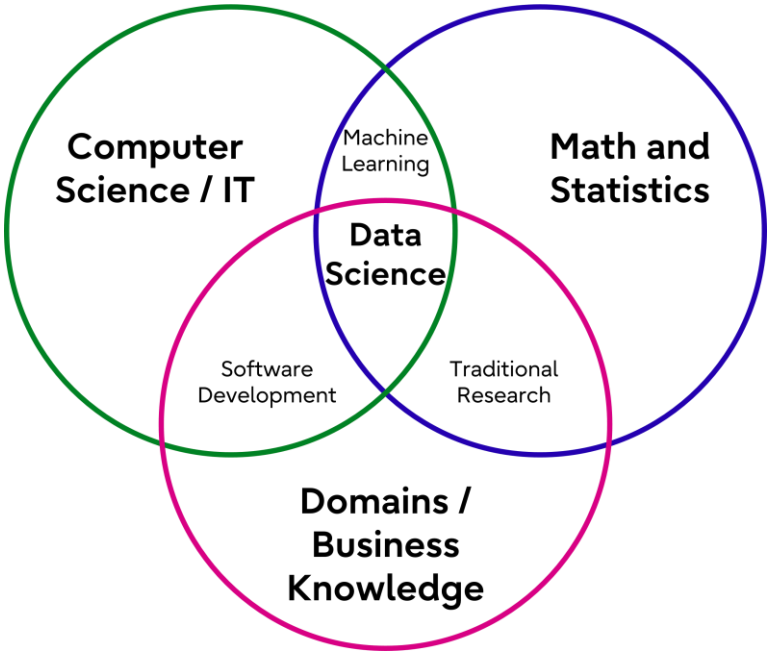
At this point in time, it is hard to picture the state of our world when more advanced types of AI come into being. However, it is clear that there is a long way to get there, as the current state of AI development is still relatively rudimentary. For those holding a negative outlook for the future of AI, this means that now is a little too soon to be worrying about the singularity, and there is still time to ensure AI safety. And for those who are optimistic about the future of AI, the fact that we have merely scratched the surface of AI development makes the future even more exciting.

Right now, complex AI tasks are handled very efficiently because of the discovery of a discipline called Machine Learning and its extension called Deep Learning.

- Machine Learning can automatically learn from and adapt to new data without being assisted by humans.
- Deep Learning enables automatic learning on huge amounts of unstructured data, such as text, images, or video. This is an extension of a set of Machine Learning algorithms called Neural Networks.

What is Data Science?

Data Science is an inter-disciplinary field of study at the intersection of Computer Science/IT, Mathematics, Statistics and Domains/Business Knowledge. It helps solve critical business problems based on scientific analysis of data using a variety of tools and techniques from Statistics, Machine Learning, Operations Research[32] and Natural Language Processing.



Data Science is leading the so-called "fourth industrial revolution", when massive amounts of data have been generated by people on a daily basis. Data Science provides a way for businesses to capitalize on this data. Our society has now become so data-driven that every major decision is a data-backed calculative move.

Data Science methods help data scientists identify patterns in this morass of data. A data scientist requires the right skills to extract, manipulate, visualize, and maintain the data to predict the occurrence of future events.

The Relationship between Data Science and AI

Artificial Intelligence and Data Science can often function interchangeably but there are some differences as well:

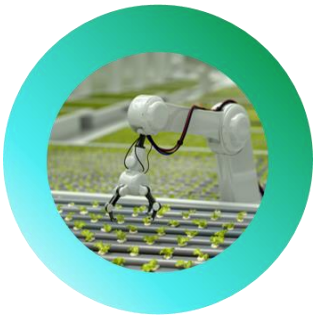
- Data Science is the analysis and study of data for extracting insights that help businesses make decisions. AI is used to build models that mimic cognition and human understanding.
- Data Science works by sourcing, cleaning and processing data to extract meaning for analytical purposes. AI combines large amounts of data through iterative processing and intelligent algorithms to help computers learn automatically.
- Data Science does not involve a high degree of scientific processing, compared with AI.
- There are many more Data Science tools than AI tools. This is because Data Science involves multiple steps for analyzing data and generating insights from it.
- Data Science is about finding hidden patterns in the data. AI is about imparting autonomy to the data model.

Machine Learning

Machine Learning is a discipline within Artificial Intelligence that is focused on using data (or interactive experience) to build intelligent systems. Machine Learning technologies are used widely in many fields, including cybersecurity, bioinformatics, natural language processing, computer vision, robotics, etc. One of the most basic Machine Learning tasks is building a classifier to automatically label objects. For example, a Machine Learning algorithm can find patterns in emails that differentiate spam from non-spam.

Some use-cases for Machine Learning:

- Over-The-Top (OTT) platforms that can stream video to any internet-connected device, like Netflix and Amazon Prime, use Machine Learning to recommend movies based on the users' past viewing data. The system constantly improves by learning from past experiences.
- In e-commerce, companies like Amazon use Machine Learning to learn users' preferences and provide product recommendations based on previous purchases and viewing history.
- Other use-cases include demand forecasting, predicting customer behavior and gauging customers' sentiments from their social media behavior.



Due to the explosion of data and the need for analyzing a large variety of structured and unstructured data sources, a novel field called Deep Learning has emerged. It encompasses methods based on the structure of the human brain. These use millions of computational units and are intended to solve complex tasks such as object detection, autonomous driving, speech understanding / translation, etc.

Concepts, Ideas and Algorithms used within Machine Learning and Deep Learning

Traditional Machine Learning Methods

- Statistical principles such as A/B testing, hypothesis testing, Law of Large Numbers, Central Limit Theorem, etc.
- Time-series models such as ARMA, ARIMA, ARCH, GARCH, etc. and concepts such as stationarity, seasonality, homoscedasticity, etc.
- Scientific libraries in Python (Scikit-learn, NTLK, Keras, Tensorflow), R, Spark/Scala, Java (DL4J), etc.
- Data structures and algorithms
- Standard probability distributions such Beta, Poisson, Exponential, Gaussian, etc.
- Statistical models on structured and unstructured data
- Supervised models such as Random Forests, Naïve Bayes, Bagging/Boosting, Perceptron, SVM, Linear/Ridge/Lasso Regression etc.
- Unsupervised models such as PCA, K-Means Clustering, Word to Vector, Restricted Boltzmann Machines, EM, Anomaly Detection, etc.
- Understanding of fundamental concepts underlying classification/regression problems, such as Empirical/Structural Risk minimization, Bias-Variance Tradeoff, Curse of Dimensionality, Overfitting, Class-imbalance problems, etc.

Deep Learning Techniques (an advanced sub-field of Machine Learning)

- Neural Networks-based training techniques, such as drop-out regularization, batch normalization, etc.
- Vanishing-gradient type of issues in training very deep models
- Recurrent Neural Networks like LSTM, GRU and their applications to time-series data
- State-of-the-art Deep Learning models, performance tuning and benchmarking
- Deep learning libraries such as Tensor-Flow, Theano, Keras, etc.
- Transfer Learning to reuse existing Deep Learning models on newer problems
- Architecting end-to-end pipelines for Deep Learning systems, such as collecting training data, noise removal, neural network design, parameter tuning, hardware configuration selection, selecting performance evaluation metrics, etc.

Operations Research (Optimization)

- Numerical optimization algorithms, such as gradient descent, back propagation, etc.
- Various types of numerical optimization, such as linear programming, quadratic programming, integer linear programming, semi-definite programming, combinatorial optimization, etc., and various methods to derive solutions to each of these types of problems.
- Mathematical modeling of real-world problems and transforming the models into those that are solvable by existing optimizers.
- Performing simulations of the optimization problems with complex objectives, constraints, and numerous parameters.
- Industry-standard optimizers, such as CPLEX, Gurobi, etc.
- Quantum / quantum-inspired technologies (e.g., Fujitsu Digital Annealer).

Natural Language Processing

- Natural Language Processing algorithms, such as HMMs, Topic Models, Latent Semantic Analysis, Named Entity Recognition, etc.
- Document Indexing and Searching frameworks, such as Apache Lucene, Elastic Search
- Text pre-processing and normalization methods, such as tokenization and POS tagging
- Models to summarize text data, such as Topic Models, Latent Semantic Indexing, Tf-Idf scoring and word embedding methods, such as Word2Vec, etc.
- Sentiment Analysis
- Named Entity Recognition methods
- Chatbot systems
- Machine Translation Models
- Recommender systems, such as collaborative filtering, content-based filtering, etc.

Fujitsu Machine Learning Technologies

At Fujitsu, there are various Machine Learning-based technologies that help solve specific types of business use-cases:

- **Semantic Search:** A search tool focused on searching based on the meaning of phrases instead of simple keywords.
- **Deep Tensor and Explainable AI:** A proprietary Machine Learning framework that works with graph-structured data to achieve reasoning and an academic basis for AI findings.
- **Akisai:** Cloud technology specifically aimed at the food and agriculture industries. It supports management of all aspects of agriculture management, such as administration, production, sales in open-field cultivation of rice and vegetables, horticulture, and animal husbandry.
- **CHORDSHIP:** A Digital Agent (chatbot) capable of highly accurate responses gained through Fujitsu's approach to Artificial Intelligence, Human-Centric AI Zinrai. It can accurately understand expressions and synonyms to provide highly accurate responses in contact center chats, narrowing down the content of the question with high precision while extracting related information from a company's knowledge bases and FAQs, then generating an appropriate response.

- **API Learning:** Enables a machine to automatically generate machine-readable format of API functionalities from API documentation, understand a variety of API descriptions, validate extracted information through automatic API validation and, finally, recommend API mashups for a specific purpose.
- **Predictive Maintenance:** Tool to predict when an asset (such as a factory part, a wind-turbine blade, etc.) needs repairing, allowing operators to schedule repairs and order parts in advance, eliminating downtime and reducing the need for high stock levels.
- **Sholark:** A suite of Natural Language Processing tools to discover links between sources of information in textual data (performs search, classification, and summarization of text).
- **Intrusive Monitoring:** Detect malware intrusions into networks within organizations leveraging Deep Tensor technology.
- **Retail Engagement Analytics:** Collects data from multiple sources, such as in-store Wi-Fi or camera feeds, and integrates it with sales data, such as POS conversion rates, or external information, such as local weather, to provide insights to enhance store operations and improve the shopper experience.
- **FAIR (Fujitsu Advanced Image Recognition):** A framework for applying Deep Learning to achieve greater automation across inspection processes. It captures all quality control findings and decisions within a defective image and compiles it into a PDF report.

Fujitsu Mobility AI Technologies

Mobility is an area of society where AI has widespread applications, with enormous potential to transform the sustainability of transport, life in densely populated regions and the experience of moving people and things from one point to another.

Digital Twin

Digital Twin Suite offers mobility-related services based on big data collected from disparate sources. Leveraging technologies and toolkits from Fujitsu such as the Digital Annealer, AI and Computer Vision, the Digital Twin platform reproduces real-world information in the digital space.

Digital Twin delivers insights to derive business value from data. For instance, leveraging data from on-board cameras, AI-powered image recognition, high-precision 3D positioning technology and analytics, the platform enables accurate estimation of the 3D position and trajectory of potential obstacles in the real world - including other vehicles, roads, pedestrians, and buildings.

CASE (Connected, Autonomous, Shared and Service, Electric)

Connected

Over the Air Updates (OTA) technology provides the capability to update the engine components of a modern car over the air without needing to bring the car into a specialized garage to apply changes. It enables remote updates to target engine controlling components. When required, components 'flash' themselves when presented with the right software package. Required updates are supplied in a secure manner, applied in a way that ensures the update will not compromise the functionality of the vehicle and its components.

Shared and Service

Aims to provide a seamless journey and great travel experience. **Spatiowl** On-Demand Transportation Services and FIDO verification match travelers with the most suitable vehicle and driver.

- Fujitsu Spatiowl delivers real-time location data analytics
- Fujitsu cloud technology creates detailed, real-time models of traffic and people-flow around cities
- FIDO identifies individuals for the sharing era
- Fujitsu biometric authentication - fingerprint, iris, palm vein, and face scanning - ensure data security.

Spatiowl is a smart mobility solution and a cloud computing framework that works with large amounts of data coming from multiple sources, such as public transportation, vehicles, and pedestrians' smartphones in urban areas through sensors. It aggregates and analyzes data across various data layers, and provides new insights from the analyzed data to identify relevant actions in the user context.

Autonomous

Autonomous driving has several challenges: engineering, regulatory, lack of industry-standardized technology and tools, and consumer trust and acceptance. The biggest challenge is the data volumes required for both training AI systems and for real-time decision making once these systems are deployed. The solution decreases data volumes and manages it in a distributed fashion, and only the relevant data is collected. With video data optimization, images are compressed and filtered. The technologies address the following challenges:

- Takes the raw data and in-vehicle components, focusing on reducing video data volumes
- Reduces the size of the data without compromising data availability or quality
- Decides what information is critical or otherwise
- Better management of huge volumes of data
- Reduces storage and associated maintenance costs
- Better availability of selected data for re-use/analysis
- Maintenance of data integrity and quality.

Electric

There are several challenges to overcome with the shift to electric cars, such as uncertainty over battery duration on journeys, choosing the most efficient traveling route, and the location of working battery recharging stations. Fujitsu's AI technologies provide more accurate battery information and create a power consumption map, and provide:

- High-dimensional statistical analysis capabilities
- The ability to learn from internal and external vehicle data, such as the driving environment
- Automatically generated battery performance models.

Conclusion

"[AI] is going to change the world more than anything in the history of mankind. More than electricity." Dr. Kai-Fu Lee (AI oracle and venture capitalist) 2018.

As text/speech translation and computer vision algorithms improve, in the near future we can expect breakthroughs in virtual assistants, autonomous vehicles, mail transportation and logistics, battery management and the production of cheaper, flexible, and durable materials.

Data is becoming an asset, with growing concern over security and ownership rights, and increasing government policies and legislation globally, specifically the current regulatory approaches of the U.S. and European Union.[26]

Technology is becoming part of our life, with the number of handheld devices rising rapidly. On average, every individual now has more than one handheld device with a microphone, video camera, recording ability and internet connection. It is augmenting our life, as information is made readily available for virtually any need.

We hope the information presented provides you with a high-level and holistic overview of AI and its possibilities. We encourage you to pick your area of AI curiosity and use the resources referenced to lift your passion from being a novice to an enthusiast. And, if you are ready to dive deeper, we will shortly publish a free whitepaper on one of the most pressing aspects of this technology - how AI can help rebuild trust in technology. If you would like to be notified when it is available, please contact **AskFujitsu**.

Authors

Shan Chidambaram

Global Head of Consulting – Data Analytics, AI, and Quantum Inspired Computing

Fujitsu America

Raghuram Nagireddy

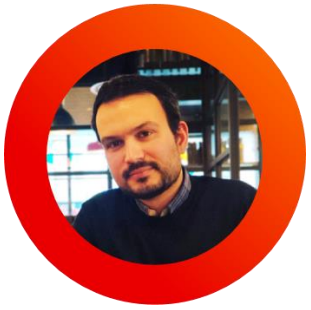
Lead Data Scientist

Fujitsu America

Biswajeet Sharma

Senior Data Engineer

Fujitsu America



José Pereira

Analytics Delivery Lead

Fujitsu Portugal

Contact

AskFujitsu

+44 (0) 1235 797711

Ref: 4084

Email: askfujitsu@uk.fujitsu.com

www.fujitsu.com