Consciousness in Autonomous Systems

Hannah Thomas, MMathCompSci, L3Harris, Portsmouth, United Kingdom
Emma Parkin, MEng, L3Harris, Portsmouth, United Kingdom

Synopsis

Standards of artificial intelligence have evolved rapidly with technology over time and today autonomous systems are expected to collaborate in more and more complex environments, such as in the field of Autonomous Surface Vehicles (ASVs).
ASVs have a requirement to operate safely in dynamically changing surroundings and interact with humans. This requires a certain level of consciousness. For the purpose of this paper, consciousness is defined as 'the state of being aware of and responsive to one's surroundings'. In order to obtain consciousness, an autonomous system must climb the knowledge pyramid of data science in real time. That is to say it must:

- Ingest (potentially huge amounts of) data from diverse sources which may contain different types of information in different formats.
- Convert each set of data into useful information about the surrounding environment and filter out noise.
- Combine this data to generate situational awareness.
- Convert this situational awareness into wisdom to evaluate the best course of action.

With over 20 years' L3Harris Technologies, Inc. understands the human-machine relationship intimately. With over 2000 days of on water testing the team has experienced first-hand how readily operators trust machines, as can be seen all around in everyday life. It is imperative to match this trust with trust-worthiness and in autonomous systems, this can only be achieved with consciousness.
L3Harris' autonomous control system, has been tested in a range of environments: missions have been conducted in daytime and night time, calm and rough seas, open water locations and busy ports with dense traffic. Each setting poses its own challenges, and autonomous systems are required to be consistently reliable. Recent advancements in technology have increased the level of consciousness achievable in systems operating today.
This paper discusses the evolution of technology that has made increasing levels of consciousness achievable today and the implications on human-machine interaction, drawing on examples from L3Harris' experience.

Keywords: Autonomy, Artificial Intelligence, Machine Learning, Human-Machine Teaming

1. Introduction: Modernising the Maritime Sector

Maritime autonomy, in particular (Autonomous Surface Vehicle) ASV technology, is not new, having been around since 1990s. What is new is the advancements in technology changing how autonomy can be achieved. There is now the ability to store and process vast amounts of data cheaply, making it possible for an autonomous system to process and interpret more data about its surrounding environment in real-time than before. In short, it is now possible to bring more consciousness into autonomous systems.

Over time, as new vessels have been introduced to the water, navigation practice has adapted to avoid collisions at sea. (Werner, 2017) notes how in the mid-1800s, the advent of steam-powered ships introduced the need for updated conventions to handle their increased manoeuvrability over sailing vessels. Today's code of practice is set out in the COLREGs (Convention on the International Regulations for Preventing Collisions at Sea), which, at the time of writing, has no specific rules for unmanned or autonomous vehicles.

While this may change in future, at present ASVs must abide by the same rules as manned vessels. ASVs must, therefore, exhibit the behaviour of a human operator. This means signalling clear navigational intent to others and exercising human-like judgement in situations with no obvious right course of action.

In situations where there is no obvious "right" answer, a compromise may have to be made and people make decisions based on personal preference, for example, a preferred type of manoeuvre. Raymond et al. (2019), proposed an argumentation framework to aid conflict resolution in maritime navigation, in which they describe such a set of preferences as a "culture". In order to act according to such a "culture" a system requires a level of consciousness beyond the minimum information necessary for simply following COLREGs. It must assess the situation according to the wider context of its own culture.

Author's Biography
Hannah Thomas: Hannah is the data science lead at L3Harris, Portchester, UK with expertise in machine learning, data analytics and big data technologies. She has developed a variety of machine learning and analytical solutions to tackle new problems across industries, including investment banking and autonomy. Hannah has a Masters degree in Mathematics and Computer Science from the University of Oxford and enjoys presenting on different aspects of her work in industry.
Emma Parkin: Emma has been a technical sales engineer at L3Harris for almost 2 years. Emma has previously been seconded to the Royal Navy, DE&S and CMRE working on a range of maritime autonomy projects. She has a Masters degree in Chemical Engineering from the University of Sheffield

 https://doi.org/10.24868/issn.2515-818X.2020.002

By contrast, it must be noted how much more accommodating of autonomous systems the infrastructures in the aviation and automotive industries are. The Civil Aviation Authority, (n.d.) describes how aviation routes are pre-defined ahead of a flight and governed through air traffic control. Road satellite navigation systems have access to unified data services to provide regular updates (Tom Tom International, 2020). Nonetheless, even in more mature environments such as these, consciousness is still required for true autonomy to facilitate reacting to unexpected situations such as obstacles appearing along the intended navigation path.

For the foreseeable future, there will likely always be a human operator somewhere in the communication loop of managing an ASV. As long as this is the case, the operator must be able to understand the system's decisions, particularly in complex scenarios. The field of Explainable AI (XAI) (Turek, 2018) has recently grown out of the Artificial Intelligence community with many researchers turning their focus to developing systems with enough consciousness to explain how and why their decisions are made. This development could see the human-machine relationship transform.

Furthermore the reaction speed of a computer is much faster than that of a human. In vehicle autonomy, this could be critical for safety in some situations. For this reason autonomous systems capable of making their own decisions and having trust from the operator to carry them out may become necessary as ASVs become more widely used.

L3Harris works to educate customers on the current state of play and the responsibility of the user when interacting with autonomy today. It remains to be seen how human-machine interaction, maritime infrastructure, regulations and resources will evolve. The level of consciousness ultimately required in autonomous vehicles is dependent on multiple factors, but it is clear that some basic level of consciousness certainly is required.

The rest of this paper discusses how machine consciousness has developed over time and its impact on how ASVs are used at sea today.

2. The Conscious Evolution

In order to understand people's expectations of technology today, how this compares to the feasible capabilities of a modern autonomy system and, consequently, how modern autonomy should interact with users, it is important to look at the historical development of consciousness in computers and how this has changed people's attitudes over time.

2.1. The Past

The notion of intelligent machines was thought of long before the research or development began. Two early examples include the books *Gulliver's Travels* by Jonathan Swift and the play *R.U.R (Rossum's Universal Robots)*. The former published in 1726, references a fictional device 'The Engine' a machine which could randomly generate literature, an illustration of which can be seen below in Figure 1. The latter released in 1920, introduced the word 'robot' to the English language, in this case robots were mass produced as labourers to serve man and ultimately revolt. Respectively written in times before computers existed, when they were industrial machines relying on punch cards *(Science & Industry Museum, 2019),* and centuries before any scientific evidence of such artificial intelligence being possible, they fed the imaginations of suggestible nations willing to entertain the idea that such a human invention could exist.

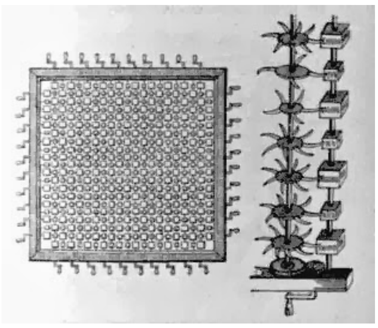

Figure 1 *Illustration of "the Engine" from Book 3 of the second edition of Jonathan Swift's Travels Into Several Remote Nations by Lemuel Gulliver (Swift 1726).*

 https://doi.org/10.24868/issn.2515-818X.2020.002

The first scientific research reasoning about the intelligence of machines was done by Alan Turing in the first half of the twentieth century. In 1936 he devised the Universal Turing Machine to understand the theoretical capabilities and constraints of a computer. After the development of machines with memory in the '40s, he published "Computing Machinery and Intelligence" (Turing, 1950), in which he concluded that, on some level, digital computers are capable of appearing, to a human, to simulate consciousness. However this could only remain a theoretical concept at the time due to the required advancement in hardware yet to come.

The rest of the century gave way to new forms of person-machine interaction, changing how the two sides communicate. Programming languages such as COBOL, FORTRAN, ELIZA (Weizenbaum, 1983) were developed in the '50s, enabling mathematicians and scientists to feed instructions through code (Computer History, n.d). In the '60s, more widely accessible graphical user interfaces were developed, allowing people to feed instructions with simple mouse clicks (Hall, n.d.). In the '70s personal computers took off and since then, computer processing power has grown exponentially, as predicted by Moore (1965) and shown in Figure 2 below. As computing resource has developed, so has machine consciousness and the level of human-machine interaction, from "expert" systems and statistical models programmed by humans to self-taught game-players (Chess (Anderson, 2017) and Go). With the advent of the World Wide Web in the '90s, internet services came into being, including the Google search engine, which made use of "intelligent" search algorithms and modern processing power to give people access to information from all over the world within seconds. With this rapid growth in accessibility and a glimpse of machine "consciousness" on a global scale, people's expectations were set to embrace conscious systems working alongside them in everyday life.

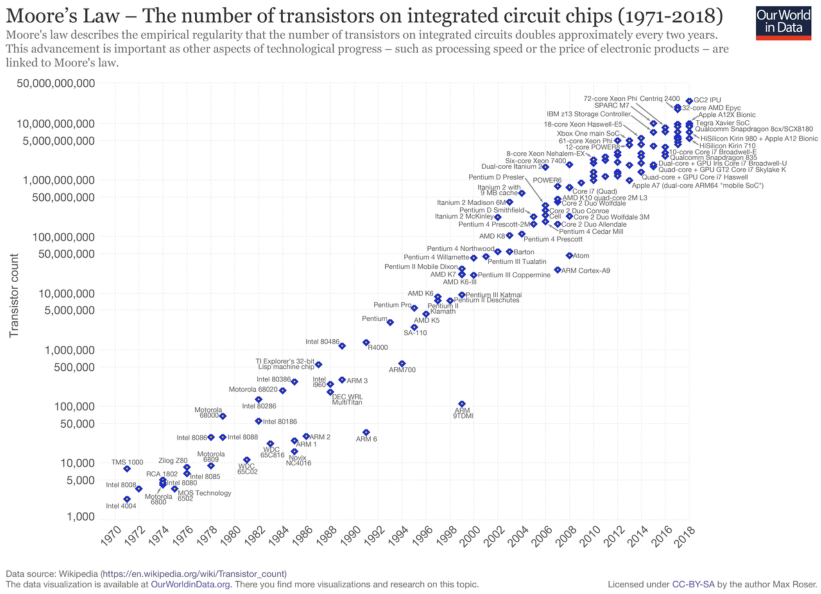

Figure 2 Processing power over time, graphed on a logarithmic vertical scale. (Wikipedia, 2020)

2.2. The Present

Swift's prediction of "The Engine" was not that far off what exists today, with systems capable of understanding human language as in speech recognition and deriving context of "similar" literature in Amazon's book recommendation service.

Recent advancements in high-performance computing and data storage units have given rise to the current age of "big data". In the present day data can be processed so cheaply that non-linear algorithms can be trained against vast data sets from high-parameter spaces, deriving all their information from data alone. This marks a step change in how researchers and developers work with computers, with fields like topological data analysis growing in popularity. Deeper exploration of "latent", real-world data spaces made up of complex, inter-related variables is now possible, allowing machines to provide more insights to aid developers engineering solutions.

 https://doi.org/10.24868/issn.2515-818X.2020.002

For example, teaching a system to identify a human face "traditionally" would include programming geometric properties such as distances between known points and positioning of the eyes (Kortli et al, 2020). These basic properties would not cater for changes like the effect of wearing glasses. Nowadays facial recognition systems are taught, from data, to recognise and ignore features like wearing glasses, change in hair styles and other "unpredictable" but irrelevant changes in appearance. Figure 3 illustrates a person using a mobile device for facial recognition in place of a more traditional password.

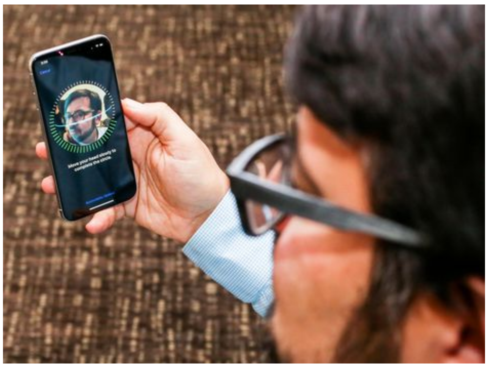

Figure 3 A representation of facial recognition technology (Tew and La, 2018)

However machine-derived hypotheses with less human input are harder for a human to explain. Now machines are able to teach themselves sophisticated, human tasks, which we would very difficult, perhaps impossible, for us to describe mathematically. An example of such a task is when we look at a chair and a bean bag, it is difficult to explain why we recognise both of these objects, which have completely different visual features, as a chair. (Goodfellow et al, 2014) describe an approach to learning such tasks, whereby machines effectively perform Turing tests on themselves.

As the continuing trend of user expectation is set for people to rely on these systems more, motivated by the need for safe operational models, researchers are turning their attention to understanding the "thinking" of these systems (Bengio et al, 2019) and (Raymond et al, 2019).

2.3. The Future

The main reason for developments in machine "consciousness" throughout the years has consistently been to aid or enhance user operations. In autonomous systems this means assisting operators safely. With modern hardware enabling the current level of artificial "consciousness" the capabilities and development of these systems are being scrutinised more and more. The next challenge the industry must tackle is maintaining user trust as these systems become available on a wider scale.

The two main ways to maintain user trust is through trustworthiness and explainability;

"First: be trustworthy. Second: provide others with good evidence that you are trustworthy." Baroness Onora O'Neill. With recent headlines about rigged elections and devices listening to your conversations (Pettijohn, 2014) it is understandable that trust is a hot topic when considering the future of autonomy and the potential for lives to be at risk.

Being ethical is a key part of building trust between the technology developer and the user. The third point identified in "6 ways tech can earn our trust" (Sutcliffe, 2017), is to "be open and show your workings" this stresses the importance of explainability. Being able to explain the inner workings of the system allows the user to understand the system and that it works well therefore improve overall trust. L3Harris has experienced this first hand as operators' trust in an autonomous control system grows as their understanding of the technology develops. Being able to explain the system can also safeguard against bias and ensure it adheres to regulatory standards (The Royal Society, 2019).

It is important to temper the rate of autonomous technology delivery with the level of trust assured in the system. Both Stephen Hawking and Elon Musk have spoken about artificial intelligence being the greatest threat to mankind. Without taking on the responsibility to be ethical in the approach there is the potential to lose control, as has been seen in the case of the Facebook/Cambridge Analytica scandal which involved the misuse of a vast amount of personal data. When it comes to adopting autonomous vehicles in broader contexts, there is potential for much wider ramifications if safety cannot be assured.

 https://doi.org/10.24868/issn.2515-818X.2020.002

"Explainable AI" is the current area of research working to address this problem from a machine point of view. However it is also important for responsibility to be taken from a user point of view. Developers today are working to understand appropriate usage of the different components of modern autonomy in order to manage expectations and establish the user role in relation to these systems.

3. The Human-Machine Relationship

As shown in section 2, human-machine interaction has progressed in both directions as machines have evolved with people over the centuries: machines have got better at communicating human-readable and useful information and people have become more computer-literate. But in the autonomy industry, the operator-AI relationship is still evolving.

(Trujillo et al, 2019) note how, with development of sophisticated vision algorithms and perception systems, the human-machine relationship in autonomy is transforming from the traditional master-servant model to one of team mates working together to achieve a common goal. They identify the necessary characteristics of good team mates as: clear communication, trust and willingness to handover tasks when necessary. They also state that, in order to work effectively alongside humans, "autonomous agents must be able to deal with conditions that may not have been foreseen during the design or every possible combination of actions considered".

This section discusses how consciousness in autonomous systems is changing the way operators are able to interact with autonomous agents as team mates.

3.1. Demonstrating Machine Consciousness

In order for clear communication, trust and effective task allocation between operators and their autonomous colleagues, the operator must be able to make sense of the AI's judgements. In order to do this, the AI must present a clear, human-interpretable view of their decision-making process. That is to say, they must demonstrate their consciousness.

This means the machine must present concise views of their internal data processing that the operator can process quickly. There is a wealth of well-established data fusion techniques (Castenado, 2013) for combining returns from different sensors into single target tracks. Being able to present the machine's internal view of the surrounding environment succinctly in this way, directing attention to important alerts, is one way of demonstrating machine consciousness to an operator.

However these alerts are often generated from underlying statistical evidence that may not be immediately obvious to a person and may make the system seem inconsistent or unreliable. Machines could demonstrate a deeper level of consciousness if they were able to explain their internal thought processes, even admitting uncertainty and providing justification for unexpected behaviour.

Addressing this gap can be a significant challenge, particularly when dealing with more complex tasks that cannot be described by a human-defined function. As an example, an image recognition system will interpret an image as a multi-dimensional array of pixels, to which it will apply a series of (non-linear) functions. Each of these functions will extract features from the original image, which are used in turn to derive some information about objects in the scene. These features provide the key to explaining how the system came to its final decision but they may be, perhaps counter-intuitively, difficult to visualise in terms of the original image. Furthermore an honest explanation would capture the statistical nature of the evidence used to generate the final decision rather than a simple binary output.

Recent research has given rise to intuitive, visual explanations of image recognition algorithms (Montavon et al. & Kohlbrenner at al.), which give indications of likely causes of specific events but still require human interpretation or further analysis for contextual reasoning. However with enough usage and consistent observations, operators may be able to build up instincts as to how the system behaves in different scenarios. It should also be noted that the quality of these explanations is generally dependent on the model used. Future work may see developers forced to focus on models that provide better explanations or development of more generalised, model-agnostic explanations.

 https://doi.org/10.24868/issn.2515-818X.2020.002

In addition to "real-time" justification of individual alerts, broader analysis demonstrating deeper levels of consciousness can be done on a larger scale off-line, revealing more insight into factors driving decisions and giving more confidence of consistent behaviour. Figure 4 shows an example of an evaluation method, proposed by the author (H. Thomas), of a complex image classification system presented for human analysis: a high-dimensional, machine representation of processed images has been mapped down to 2 dimensions, demonstrating how the system is separating out the different classes. It shows which examples lie at the boundary edges, representing "grey" areas of uncertainty, those which are easily classified and those which are poorly understood by the system.

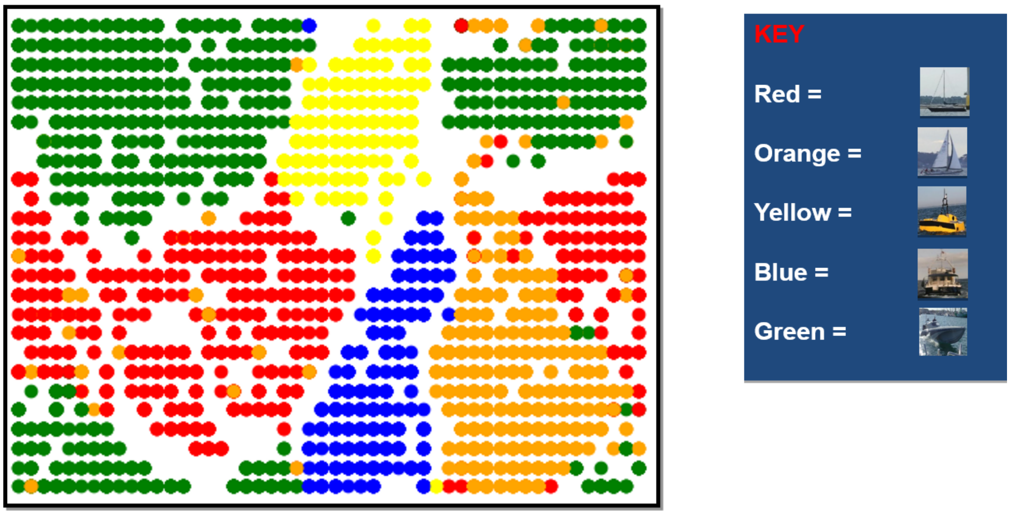

Figure 4 - A visualisation of an image recognition system model, demonstrating how well the individual classes have been separated by the model. The map wraps around from top-bottom and left-right.

3.2. Human Feedback and Continuous Learning

People continuously receive feedback from constant interactions with their environment which develop their views and behaviours. For example, children learn language through constant communications with parents. Initially their capability is limited but with constant learning their vocabulary quickly expands and grasp of grammatical rules strengthens.

Developing an equivalent feedback process for machine learning is an active area of research known as reinforcement learning (Kiran et al, 2020 & Imanberdiyev, 2016), which has started to see success in autonomy (Karpathy, 2020). Such a process enables operators to continuously teach autonomous systems how to respond to new scenarios, increasing their vocabulary for handling the variety of situations faced in real-world operations. It also encourages autonomous behaviour to be more human-like and thus more predictable for their human counterparts. For example, an autonomous navigation system that responds to every millisecond update with a change in direction can appear very indecisive and confusing to a person. This is potentially dangerous on the water where clear signals of intent between actors are required.

The feedback process can in fact work both ways. Whilst the operator teaches the machine their way of thinking, they are simultaneously trained in the machine's way of "thinking". They are also forced to evaluate their own decision-making processes. This regular interaction and exposure to machine consciousness could potentially be a good model for human-machine team training.

3.3. 'Required' Levels of Consciousness

As autonomous team mates are emerging, there is still much research to be done on how much machine consciousness is useful and how much should be communicated to the operator. *(Chen et al, 2017)* propose a 3-tiered situational awareness model to organise levels of autonomous communication. Level 1 simply provides

 https://doi.org/10.24868/issn.2515-818X.2020.002

current state and goals, intentions and plans. Level 2 provides the agent's reasoning process before making decisions. Level 3 provides projections of future outcomes, uncertainty and potential limitations, likelihood of success and performance history.

Using their model, (Chen et al) carried out experiments to analyse the effect of machine transparency on human trust. They found communications of uncertainty had differing outcomes depending on the role of the agent: when acting as operator decision aids, they were granted more trust; when acting as autonomous team members, they were not granted more trust.

These results could be a symptom of users not being used to working with autonomy in this way. Historically people have been used to relying on computers to perform deterministic tasks, like numerical calculations. But now they are being taught to perform more nuanced, non-deterministic tasks, requiring human-like consciousness.

Whilst these technologies are still developing, it might seem safer to demonstrate the internal machine thought process, allowing the user to understand and make corrections as necessary. In this case, perhaps autonomous agents are better used as decision aids or restricted to tasks in more controlled environments whilst still in training, until they reach a level of consciousness comparable to human judgement.

Ultimately plenty more experimentation of human-autonomy interactions is still required to understand effective operating models but it is likely both sides will need to adapt to each other. If it is not practical for users to be exposed to the underlying uncertainty of machine decision-making, the same leniency we have for human decision-making will need to be extended to machines. Equally research is required to better understand how to evaluate machine performance on such uncertain tasks and at what point they should be granted the same trust as humans.

4. Conclusions

It can be seen that as consciousness has matured in machines, people's expectations of their "human-like" qualities has increased. However it is still an open question for the autonomy industry as to what level of consciousness is achievable, required, and moreover desirable. Does operator behaviour need to evolve or the sophistication of the system or, more likely, both?

As history has taught us, there is a middle ground to be reached. This can be seen from the industrial revolution in the late 1700s to early 1800s where the way goods were produced dramatically changed. Current ways of working and regulations, across all industries and operations, were defined for people and assume people-to-people interactions. However, a conscious system will internally evaluate situations in a different way to a person (even if the same outcome is reached). In the same way machines have evolved to explain internal processes to the user, and people have become more computer literate over time, a centre ground needs to be reached for people to work harmoniously with the current generation of technology. In ASVs, this may mean increased transparency from the AI and increased tolerance of AI uncertainty from the operator, treating the autonomous system more like a human peer.

As systems develop more consciousness and appear to behave more like humans, people generalise from a small set of observations and expect to observe this to apply more broadly. However, the mere fact that a system appears to behave like a human in certain scenarios does not necessarily mean it has learnt to generalise this behaviour: a different scenario with different parameters could lead to a totally different outcome. Perhaps the future of explainable AI should focus on addressing this gap between user expectation and realistic AI functions.

Conscious autonomy systems that can simulate human thought processes that we cannot define, on a broader scale, necessitate significant, regular training and correcting from humans. This is the only way to capture the variety of scenarios in which the system is required to perform. This dictates a transformation in operator and machine roles, where people are required to actively share knowledge with their autonomous team mates. If machines were able to equally share their thought processes, this would give the user a better chance of correcting or ratifying their behaviour and, ultimately, lead to a more intimately bonded team. Perhaps developers should turn their attention to providing visualisations that provide the user a deeper understanding of the underlying function.

Increasing levels of consciousness have changed the playing field for autonomous systems, allowing machines to perform more and more active lookout and maintenance functions, but the operating model of a human-machine team is yet to be optimised. As operators begin to work more closely with conscious AI agents, the roles of the two parties will naturally evolve and we will likely see adaptations on both sides.

References

Atanasoff-Berry Computer. (n.d.). Retrieved from Computer History: https://www.computerhistory.org/revolution/birth-of-the-computer/4/99

Anderson, M. (2017, May 11). Twenty years on from Deep Blue vs Kasparov: how a chess match started the big data revolution. Retrieved from the conversation: www.theconversation.com/twenty-years-on-from-deep-blue-vs-kasparov-how-a-chess-match-started-the-big-data-revolution-76882

Castenado, F. (2013). A Review of Data Fusion Techniques. The Scientific Qorld Journal

Chen, J. Y. C., Wright, J. L., Barnes, M., Stowers, K. (2017): Situation awareness-based agent transparency for human-autonomy teaming effectiveness. DOI: 10.1117/12.2263194

Civil Aviation Authority. (n.d.). Retrieved from How air traffic control works: www.caa.co.uk/Consumers/Guide-to-aviation/How-air-traffic-control-works/>

Farag, A. A. (2019). Geometric features extraction. Cambridge University Press

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D,. Ozair, S., Courville, A,. Bengio, Y. (2014). Generative Adversarial Nets. Neural Information Processing Systems Conference.

Grace Murray Hopper. (n.d.). Retrieved from Computer History: https://computerhistory.org/profile/grace-murray-hopper

Hall, M. (n.d.). Douglas Engelbart. Retrieved from Britannica: https://www.britannica.com/biography/Douglas-Engelbart

Imanberdiyev, N., Fu, C., Kayacan, E., Chen, I (2016). Autonomous navigation of UAV by using real-time model-based reinforcement learning. 14th International Conference on Control, Automation, Robotics and Vision (ICARCV).

Kapko, M. (n.d.). History of Apple and Microsoft, 4 Decades of Peaks and Valleys. Retrieved from CIO: https://www.cio.com/article/2989667/history-of-apple-and-microsoft-4-decades-of-peaks-and-valleys.html

Karpathy, A. (Feb 2020): AI for Full Self-Driving, The Scaled Machine Learning Conference 2020 (Presentation available here: https://info.matroid.com/scaledml-media-archive-2020)

Kiran, B. R., Sobh, I., Talpaert, V., Mannion, P., Sallab, A. A. A., Yogamani, S., Pérez, P (Feb 2020): Deep Reinforcement Learning for Autonomous Driving: A Survey

Kohlbrenner, M., Bauer, A., Nakahima, S., Binder, A., Samek, W. & Lapuschkin, S. (Jul 2020): Towards Best Practice in Explaining Neural Network Decisions with LRP

Kortli Y, Jridi M, Falou AA, Atri M (2020) Face Recognition Systems: A Survey. Sensors (Basel). doi:10.3390/s20020342

Lemelson-MIT Program. (n.d.). ENIAC (Electronic Numerical Integrator And Computer). Retrieved from Lemelson-MIT Program: https://lemelson.mit.edu/resources/john-mauchly-j-presper-eckert

McCarthy, J. (2007). What Is Artificial Intelligence?

Montavon, G., Samek, W., Mueller, K-R. (2017): Methods for interpreting and understanding deep neural networks

Moore, G. (1965). Cramming more components onto integrated circuits. Electronics, Volume 38, Number 8, 114.

O'Reilly (2020) https://www.oreilly.com/library/view/java-deep-learning/9781788997454/60579068-af4b-4bbf-83f1-e988fbe3b226.xhtml

Pettijohn, N. (2014, September 25). Of Course Your Phone Is Listening To You. Retrieved from Forbes: www.forbes.com/sites/nathanpettijohn/2019/09/03/of-course-your-phone-is-listening-to-you/?fireglass_rsn=true#19e012066a3f

PROGRAMMING PATTERNS: THE STORY OF THE JACQUARD LOOM. (2019, June 25). Retrieved from Science and Industry Museum: https://www.scienceandindustrymuseum.org.uk/objects-and-stories/jacquard-loom?fireglass_rsn=true

Raymond A., Gunes H., & Prorok A.: "Culture-Based Explainable Human-Agent Deconfliction" arXiv:1911.10098v1 [cs.MA] 22 Nov 2019

Tew, S., La, L. (2018, August 22). 10 best phones with facial recognition: iPhone X, Note 9, LG G7 and more. Retrieved from CNET: https://www.cnet.com/news/10-best-phones-with-facial-recognition-iphone-x-note-9-galaxy-s9-lg-g7/

Sutcliffe, H. (2017, September 1). 6 ways tech can earn our trust. Retrieved from World Economic Forum: https://www.weforum.org/agenda/2017/11/trustworthy-tech-fourth-industrial-revolution/?fireglass_rsn=true

Swift, J. (1726). Gulliver's Travels. London: Benjamin Motte.

The Royal Society (2019, November). Explainable AI: The Basics. United Kingdom: The Royal Society.

Tom Tom International. (2020). eal-Time Map Data. Retrieved from Tom Tom: https://www.tomtom.com/products/real-time-map-data/

Trujillo, A.C., Gregory, I. M., Ackerman, K. A. (2019) Evolving Relationship between Humans and Machines

Turek, D. M. (2018). *Explainable Artificial Intelligence (XAI)*. Retrieved from Defense Advanced Research Projects Agency: www.darpa.mil/program/explainable-artificial-intelligence

Turing, A. (1950). Computing Machinery and Intelligence. Mind, 433-460

Weizenbaum, J. (1983). ELIZA — a computer program for the study of natural language communication between man and machine. Communications of the ACM.

Capt. Jeff Werner (2017) www.allatsea.net/the-history-of-the-rule-of-the-road/

Wikipedia (2020) https://en.wikipedia.org/wiki/Moore%27s_law

Yoshua Bengio, T. D. (2019, January 30). A Meta-Transfer Objective for Learning to Disentangle Causal Mechanisms. arXiv:1901.10912

 https://doi.org/10.24868/issn.2515-818X.2020.002