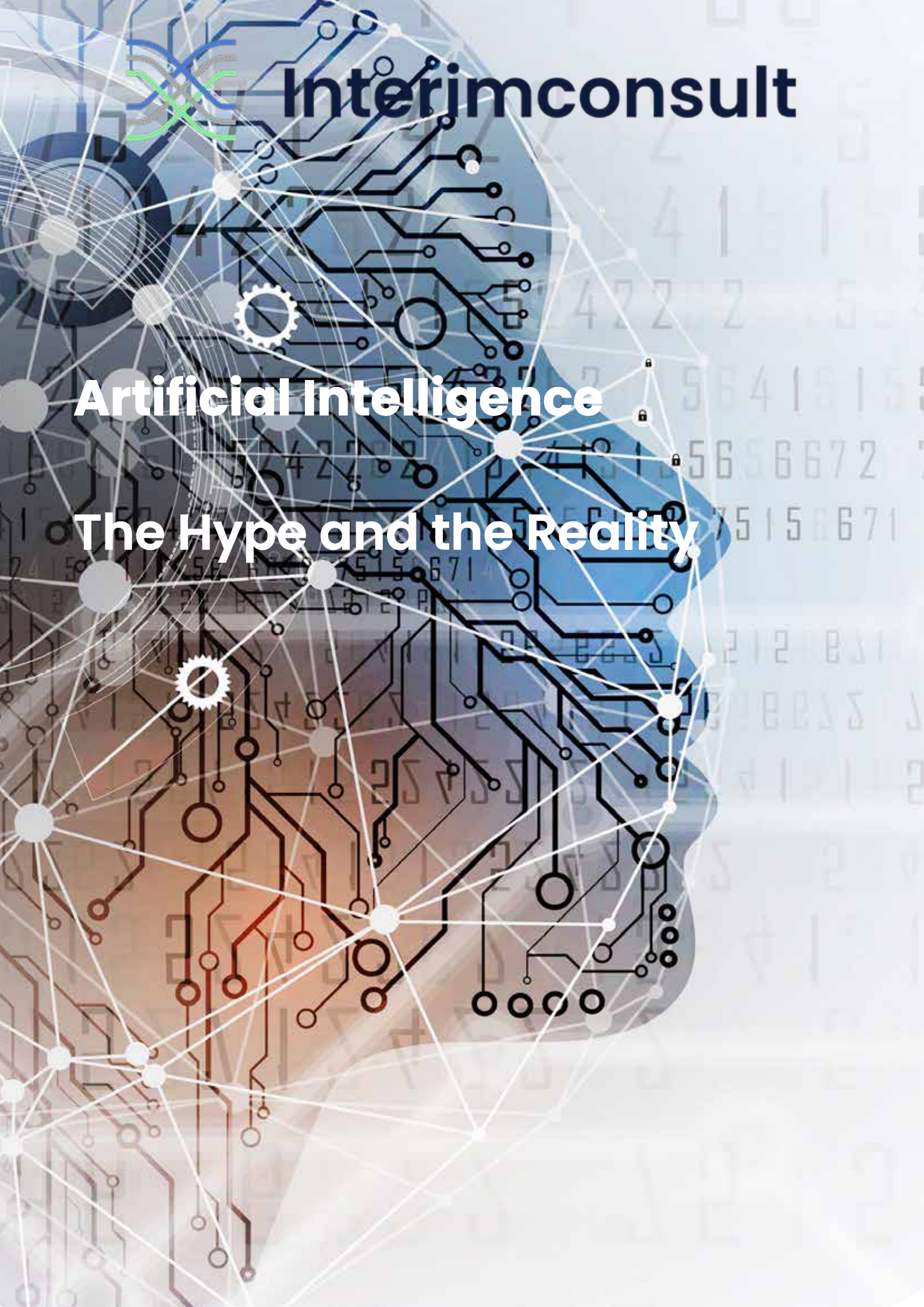

Interimconsult
Artificial Intelligence
The Hype and the Reality

Artificial Intelligence – The Hype and The Reality

There is much misinformed hype about apocalyptic scenarios where Artificial Intelligence (AI) becomes more intelligent than humans. But rest assured, whilst we need to accept the fact that the development of new applications made possible by AI will be as disruptive as the industrial revolution, and may be exploited by bad actors, the models that underpin AI, including deep learning using neural networks, will never make machines more intelligent than humans.

It is simply not possible to take the step from what we know as AI today, to what we would describe as 'general artificial intelligence' - the human capacity for creativity, innovation and what is widely considered as just being "common sense".

Claims to the contrary are at best misinformed and at worst deliberately misleading.

Nonetheless, innovations in AI do pose significant threats to industries, businesses and society, with the potential for disruption amplified by enabling innovations in quantum computing, global network connectivity, and an exponential increase in data availability.

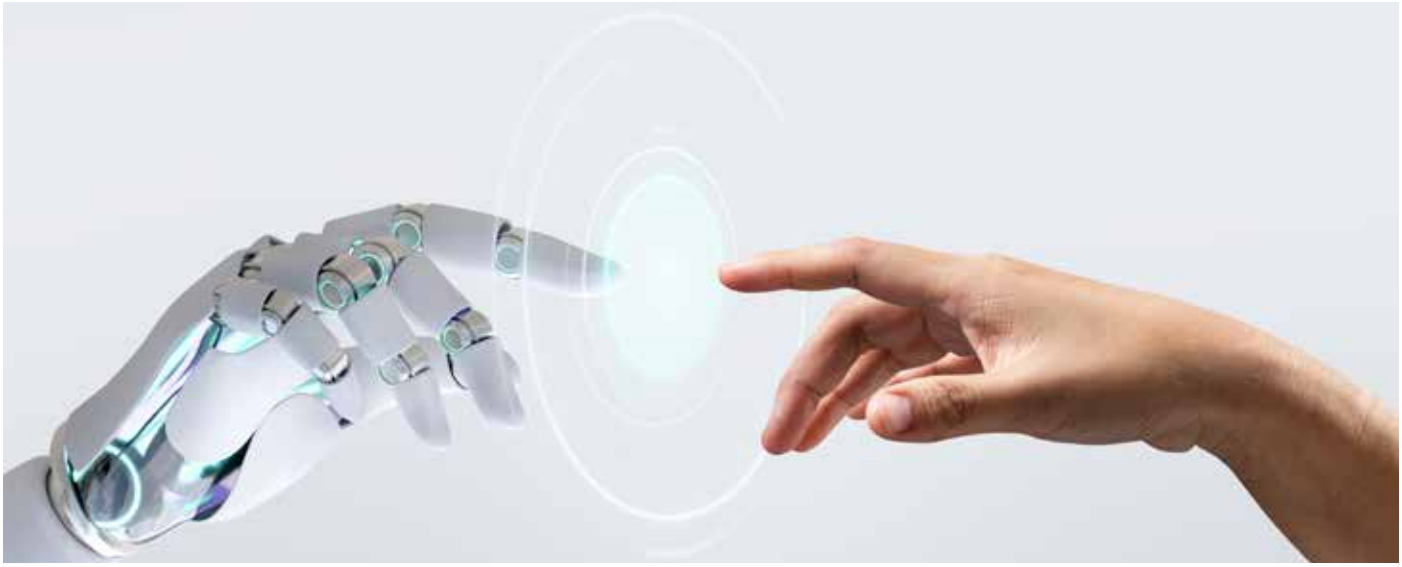
This combination means that the impact of AI could be as significant as three of the greatest global changes over past centuries, the industrial revolution, the development of nuclear technology, and the growth of global network connectivity.

There is in fact a strong analogy with nuclear technology. On the one hand, we have weapons of mass destruction, and on the other we have a limitless green energy source. What matters moving forward is how the technology is used and regulated. The same can be said for AI – how AI is used and regulated creates both opportunities and threats, not the availability of AI itself.

In this scenario, one of the key responsibilities for people and businesses working in the artificial intelligence arena is to demystify what AI can, and cannot, do. The nature of the subject is complex, but more energy needs to be devoted to developing accurate narratives about AI so that businesses and people can understand the likely impact of what is fast becoming the most disruptive technology development in recent decades.

As with the industrial revolution, despite the threats there will be many positive developments and new types of jobs will be created. AI systems are particularly good at pattern recognition, feature detection, optimisation, and forecasting - so there are huge opportunities and benefits in many areas such as health, transport, infrastructure, energy, and the battle against climate change.

Every sector will be disrupted so the recommendation is to learn more - but there's no need to worry about the world being ruled by AI.



The Nature of Artificial Intelligence

The use of the term artificial intelligence and its association with human intelligence, instigated by Alan Turing, has unintentionally led to a colossal misunderstanding about the nature and capability of artificial intelligence.

Human intelligence is developed in a complex world through day-to-day experiences that cannot be captured in data alone. Access to a wealth of data, even when updated in real-time, does not, and cannot, create the capacity for creativity, innovation and what we consider "common sense".

As an example, one of the biggest threats of AI, impersonation, and deception, is most successful when conducted remotely. Written, video and verbal communication, generated by AI, can create the perceived existence of a real person behind the screen – the criteria for passing the Turing Test.

This level of superficial engagement does not compare to the immediate feelings and emotions generated when meeting somebody face-to-face in a new environment – where an immediate human response and assessment, enhanced by the ability to reflect in slower time, is based on thousands of years of sapiens evolution.

Like the outcomes of the industrial revolution, AI's impact will come in the form of people outsourcing their activities to machines – not just at work but also at home.

This will be the catalyst for huge disruption since the AI developments in vision, speech, pattern analysis, inference and inductive thinking mean there are a huge number of activities that can be outsourced.

Some of these activities will be simple and some will be complex. Those that are most complex will lend themselves to augmentation of human capability or user-in-the-loop engagement.

The misunderstanding about the potential impact and applications of AI can be addressed, in part, by a definition that reinforces the distinction of artificial intelligence from human intelligence.

Defining Artificial Intelligence

One of the clearest sets of terms defining AI was adopted in May 2019 in a suite of legal definitions from the OECD (OECD/Legal/0449) based on a recommendation from its Council on Artificial Intelligence.

1. **An Artificial Intelligence (AI) System** is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.

2. **AI system lifecycle phases** involve: (i) 'design, data, and models'; which is a context-dependent sequence encompassing planning and design, data collection and processing, as well as model building; (ii) 'verification and validation'; (iii) 'deployment'; and (iv) 'operation and monitoring'. These phases often take place in an iterative manner and are not necessarily sequential. The decision to retire an AI system from operation may occur at any point during the operation and monitoring phase.

3. **AI knowledge** : AI knowledge refers to the skills and resources, such as data, code, algorithms, models, research, know-how, training programmes, governance, processes and best practices, required to understand and participate in the AI system lifecycle.

4. **AI actors**: AI actors are those who play an active role in the AI system lifecycle, including organisations and individuals that deploy or operate AI.

5. **AI Stakeholders**: Stakeholders encompass all organisations and individuals involved in, or affected by, AI systems, directly or indirectly. AI actors are a subset of stakeholders.

The OECD Artificial Intelligence Values-Based Principles

Importantly, the recommendation identifies five complementary values-based principles for the responsible stewardship of trustworthy AI. It calls on all AI actors to promote and implement them.

- Inclusive growth, sustainable development, and well-being.
- Human-centred values and fairness.
- Transparency and explainability.
- Robustness, security, and safety.
- Accountability.

Importantly, with the accelerated development of AI, people will need to know when, and when not, they are working with or exposed to AI.

The OECD definition and set of principles meets the test of responsibility for developing accurate narratives about AI that both businesses and people can understand. A common understanding can then be used as a basis for assessing the potential impacts of AI upon businesses, people and society.

It is clear from the definitions that AI itself and AI knowledge is not equivalent to general human intelligence and human knowledge. AI intelligence and knowledge is restricted to a machine-based construct. System functionality such as forecasting, classifying, predicting and decision support are rooted in a system lifecycle, not dissimilar to that of other disruptive technologies.

It will be the AI actors and stakeholders that will determine the opportunities and threats associated with AI and its application.

Artificial Intelligence - The Issues

Even with the best intentions, AI applications are inherently susceptible to a number of issues through what might best be describes as 'errors of commission'.

Looking beyond the computational challenges associated with implementing if-then logic, content tagging and vocabulary parsing that can be readily made available under the banner of AI, the development of AI systems is constrained by the data it uses to 'learn'. And this leads to many challenges including the selection and validity of data, its use for training and testing, and most importantly its coverage.

Many things can be represented in data, but many cannot.

Learned Bias

Bias can be introduced when AI systems are trained on datasets that are (unintentionally) biased, this type of supervised learning will increase the propensity for false positive or false negative results or, in the case of forecasting, 'over-fitting' or meaningless correlations.

The validation, selection and data variable encoding are key factors in determining whether bias becomes embedded in the AI system. The data must be from a source that can be relied on for accuracy and context and the selection of data samples used for the different phases of 'learning; and 'testing' must not contain hidden biases.

At a configuration level, the data encoding will have a significant impact on the AI system's performance. Data variables should be encoded carefully to retain key features since inappropriate encoding can lead to the most important features in the data being lost in the noise of a large dataset.

Amazon Recruitment Tool Bias

In 2018, Amazon stopped using a hiring algorithm after finding it favoured applicants based on words like "executed" or "captured" that were more commonly found on men's resumes, and downgraded references to all-women colleges and inclusion of the word "women's".

"Amazon's computer models were trained to vet applicants by observing patterns in resumes submitted to the company over a 10-year period. Most came from men, a reflection of male dominance across the tech industry. In effect, Amazon's system taught itself that male candidates were preferable".

Reuters - Amazon scraps secret AI recruiting tool that showed bias against women (2018)

Gender Recognition

In 2018, Jay Buolamwini, and Timnit Gebru in their ground-breaking MIT research paper audited commercial gender classification systems produced by Microsoft, IBM, and Megvii's Face++. The results showed that the rate of misclassification for darker-skinned women was around thirty-five times higher than for white men.

"We evaluate 3 commercial gender classification systems using our dataset and show that darker-skinned females are the most misclassified group (with error rates of up to 34.7%). The maximum error rate for lighter-skinned males is 0.8%. The substantial disparities in the accuracy of classifying darker females, lighter females, darker males, and lighter males in gender classification systems require urgent attention if commercial companies are to build genuinely fair, transparent and accountable facial analysis algorithms".

Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification

Jay Buolamwini, and Timnit Gebru (2018)

Maintenance

Deploying AI systems that are designed for a specific purpose must always be updated with the latest data when the context changes. AI systems learn by 'looking back or at' data at a specific point in time so as context changes, the AI systems needs to be updated. Failure to do so can lead to the AI system losing its fitness for purpose.

AI systems that are developed in data-light environments or are subject to seasonality or changing patterns over time are susceptible to this type of error. It is most relevant to one of the most common AI applications - forecasting - where future predictions are made by looking back at past data.

The statement 'all forecasts are wrong' is an over-simplification since what matters is to what degree the forecast is wrong. If the data used to build a forecast is up to date and relevant, the forecast will be more accurate – if the data is old or becomes less relevant than newly available data, the forecast will be less accurate.

Long-range weather forecasting is a prime example. The UK Met Office are already developing systems combining physics-based and AI approaches to produce data that can be used to train fast AI-based emulators. Climate change means the data is not available to forecast its impact, so the data used to create an AI-based forecast is generated from other AI-based forecasts. These will no doubt contain forecasting errors that will be compounded but will also not include the impact of unpredictable events such as volcanic eruptions that are yet to occur.

"The forecast is based on the key drivers of the global climate, but it doesn't include unpredictable events such as large volcanic eruptions, which would cause a temporary cooling".

The UK Met Office Temperature Forecast for 2023

So the validity of an immediate climate-change forecast is limited and cannot be relied upon without further updated should there be a significant volcanic eruption. In 2022, 33 eruptions started and 80 were active across 28 countries but none were major. In 2021 there were also 33 eruptions, with 76 active volcanoes, but this time there were 3 with a 'high explosivity index' (Smithsonian Institution – Global Volcanism Program).

In essence, forecasts can only ever be as good as the data they are built upon. If key data that influences outcomes is not readily available - such as the timing, location and size of volcanic eruptions - their atmospheric cooling effect will not be included in a long range weather forecasting model.

This does not undermine the value of the forecasting activity, but emphasises the importance of ensuring that data is maintained to reflect the current context.

Explainability and Ambiguity

In deep learning, using large scale neural networks means it is not possible to explain why the inputs to a system lead to the system outputs - so it is not possible to describe 'what's in the AI box' and why and AI system is making a particular prediction or decision. Then, as systems adapt over time, the audit of what is changing 'in the AI box' becomes more complex.

In the case of decision support, where a user is in the loop, this low level of transparency can undermine trust in the AI system's recommendations. Wider intelligence not captured within the system due to limited data or by design means the user themselves will have to learn, over time, when the results can be trusted and to what degree.

However, in 'closed loop' fully autonomous deep-learning systems without the need for human intervention, and with no user to interpret or moderate the AI system decision, the nature of AI can lead to what might be inexplicable or misleading results. The reason is that AI systems are developed based on the input data provided to them and learning algorithms that are used to achieve specific outputs.

Explaining 'what is in the AI box' and understanding ambiguity and its implications is not as complicated as it may seem, with three examples given below.

- **Identification** - face recognition is based on pattern matching between biometrics extracted from digital images (e.g. the relative positions, differences, and sizes of facial features, etc.) and their comparison with datasets of people's actual facial biometrics. The goal is to match the biometric data and the outputs is usually presented in the form of a 'match probability percentage'. Contrary to common beliefs, identification is never 100% certain an already proven case has been demonstrated by printing 3D masks that depict faces with a high degree of authenticity.

- **Classification** – classification - sometimes called categorisation - is based on making a 'hard' binary decision (i.e. yes/no) as to whether a particular pattern 'fits' a set of categories (classes). A lack of certainty, similar to that in identification systems, is often the root cause of classification errors. Everyday examples of classification errors in AI - such as false positive or false negative errors - include spam email identification, online content tagging, and satellite navigation route choice.

- **Content feeds** - the customisation of social media content is based on the feed items being tagged as of a particular type – the tags may be extracted automatically based on keyword analysis. For example, sports news may have several tags that identify the item as sport, the type of sport, and the type of content (e.g. image, article, video, etc.). Tailored feeds are generated depending on the user click frequency of content with the same tags – with the clicks used to rank tags as a measure of user interest. The most highly ranked tags are then used to tailor the social media feed. Further clicks on the tailored feed will then refine and reinforce the choice of items presented to social media. Users of Instagram, Twitter, LinkedIn, Facebook and similar feeds will already have seen this reinforcement effect – if you are interested in AI, your feed will include a high proportion of items related to AI.

Ethics, Governance and Regulation

There can be no doubt that AI is already being exploited for unethical use. So rather than just treat AI as a disruptive technology phenomenon, the reach of its applications demands ethical guidance, strong governance, and workable regulation are established sooner rather than later.

Governance

Governance is a particular challenge for businesses and their supply chains where the increasing use of AI becomes embedded into day-to-day operations. AI systems may inadvertently be hidden within bought-in products and services, or it may be developed in-house as part of a business' core activities. Both scenarios can lead to problems associated with hidden bias, impact assessment, data security, and explainability.

In 2023, the Institute of Directors' Science, Innovation and Technology Expert Advisory Group published a paper to help members to understand the impact of AI and how to go about its governance.

"AI can amplify existing bias in human decisions and safeguards are needed to prevent this and its impact on perpetuating bias in existing culture".

AI in the Boardroom - The essential questions for your next board meeting - IoD Business Paper

The core principles of ensuring transparency and explainability are embedded throughout the IoD paper, which also highlights the need for defining board accountabilities, measurement of AI use, the testing for bias and the need for regular audit and review.

Regulation

The publicity around AI has led to some recent calls to pause development. In reality, this is not possible. The foundations of AI were being developed in the mid 20th century and their development and use was an inevitability - now enabled by increased and low cost access to computer processing power, memory size, and an explosion of data. The algorithms are the same, the application methodologies and constraints are the same, but the implementation barriers have been removed.

This situation demands regulation but the pervasive nature of AI will make this a complex task. Particular attention will be needed to look at the balance between generic technology regulation, with specific market regulation across international boundaries.

"Regulate content now to avert risks, says ChatGPT boss Sam Altman"

Artificial intelligence could be used to influence elections, manipulate financial markets and allow criminals to create "counterfeit people", American lawmakers have been warned at a hearing to discuss the risks and benefits of the technology.

Sam Altman, chief executive of OpenAI, whose ChatGPT tool has attracted widespread attention since its release last year, told Congress that government regulation would be "critical to mitigate the risks of increasingly powerful" AI systems, and acknowledged that the technology could do "significant harm."

He insisted that ultimately generative AI would "address some of humanity's biggest challenges, such as climate change and curing cancer", but conceded that the technology posed significant risks, accelerating the spread of disinformation and upending the jobs market.

The Times - 17th May 2023

Artificial Intelligence - The Threats

The hype around AI has led to some recent calls to pause development but the already ubiquitous availability of the technology components, international market forces and geopolitics means this is not possible. In these circumstances It is useful to look at how AI might be misused deliberately, or how it might fail due to design, deployment and implementation issues.

Many of the forms of misuse are already common and known to organisations and people already - but their rapacity will increase.

Misrepresentation and fake content creation

This will typically involve manipulating images or videos and speech generation to create purposefully misleading information - but on a mass scale. The rate at which this type of content can be automatically generated can lead to floods of what might appear to be authentic content that is rapidly spread across social media.

The illusion of authentication, and validation, can be created by the speed and volume of repetition of the same message across multiple platforms. This is not a new phenomenon.

"Falsehood flies, and truth comes limping after it, so that when men come to be undeceived, it is too late; the jest is over, and the tale hath had its effect".

Jonathan Swift – Author, Essayist - The Examiner (1710)

Social media has already highlighted the risk of this mass influencing phenomenon and its tangible impact. Actions and decisions triggered by social media are often quoted regarding the results of elections and referendums, content published by the Covid 19 anti-vax movement, and climate-change denial. In the same way that people have learned to filter some fake news and assess the veracity of the content they are exposed to, the ubiquitous capability for AI-generated content creation will make this much harder in the future - especially for those that are most vulnerable.

Facebook and Galactica

In November 2023, Facebook launched Galactica, which was taken down after 3 days having been found to be generating text with significant errors – but with a tone of authority. Having been tested extensively prior to its release, it was not possible to explain why Galactica generated the outputs it did.

"A fundamental problem with Galactica is that it is not able to distinguish truth from falsehood, a basic requirement for a language model designed to generate scientific text. People found that it made up fake papers (sometimes attributing them to real authors), and generated wiki articles about the history of bears in space. It's easy to spot fiction when it involves space bears, but harder with a subject users may not know much about".

Will Douglas Heaven - MIT Technology Review

Impersonation

AI will increase opportunities for impersonation through the creation of convincing counterfeit people and businesses. Much of the impersonation is currently committed by people responding remotely over the internet and developing conversations to form relationships based on the target victim's responses, The need for a human criminal to be involved is a limiting factor when it comes to contacting their target population of victims.

But over time, more sophisticated chatbots using generative AI for speech and written communications, its use of publicly available personal metadata, data theft, and natural language processing has the potential to break this bottleneck and increase the criminal's bandwidth for fraud. Even now, criminals can already increase their capability to impersonate people by using AI to paint a metadata picture from breaches of GDPR and publicly available data on individuals, including biometric data through manipulation of images and videos.

Like fake news, this is not a new occurrence. Online fraud is now the highest category of fraud in the UK with Ofcom reporting 87% of adults in the UK had come across content they suspected to be a scam or fraud, with nearly half of them having been drawn in.

T*he proportion of fraud incidents that were cyber-related increased to 61% from 53% in the year ending March 2020*

ONS Crime Survey Census 2021

Cyber Security

Cyber attacks are increasing in frequency and sophistication, with a constant battle between those trying to attack, and those trying to defend. Many of the tools and techniques for both roles are similar in terms of technology – threat detection is often related to vulnerability identification.

AI has the potential to change the balance between threat detection and attack. Sophisticated tools that detect anomalies and patterns across large and complex networks are used to identify and take countermeasures enabled by AI systems.

However, the equivalent scenario also applies to cyber attacks – AI can be used to used to counteract threat protection by burying anomalies in high volumes of network traffic and generating decoy patterns that consume cyber-threat detection capacity.

More complex AI 'systems of systems' approaches can be used for defence or attack. The technological challenge for responsible users of AI will be to ensure the security measures stay ahead of the cyber threats, especially those associated with critical national infrastructure and military capabilities.

Manipulation and Coercion

It is an unfortunate fact that there are a significant minority of people that will use whatever means is available to harm others, and online content and social media engagement provides many such opportunities.

Part of the challenge can be addressed through regulation to increase the pressure for companies to moderate their platform content by changing laws to help protect adults and children online.

"Illegal content that platforms will need to remove includes: child sexual abuse, controlling or coercive behaviour, extreme sexual violence, fraud, hate crime, inciting violence, illegal immigration and people smuggling, promoting or facilitating suicide, promoting self harm, revenge porn, selling illegal drugs or weapons, sexual exploitation and terrorism".

A Guide to the Online Safety Bill (2022)

However, the autonomous algorithms that increase the frequency of content presented to users based on tagging clicks and browsing history can lead to devastating effects. At the coroner's inquest into the death of 14-year-old Molly Russell, the coroner concluded that Molly died from "an act of self-harm while suffering from depression and the negative effects of online content".

The platforms operated in such a way using algorithms as to result, in some circumstances, of binge periods of images, video clips and text some of which were selected and provided without Molly requesting them.

These binge periods, if involving this content, are likely to have had a negative effect on Molly. Some of this content romanticised acts of self-harm by young people on themselves. Other content sought to isolate and discourage discussion with those who may have been able to help.

In some cases, the content was particularly graphic, tending to portray self-harm and suicide as an inevitable consequence of a condition that could not be recovered from.

"The sites normalised her condition, focusing on a limited and irrational view without any counterbalance of normality

"It is likely that the above material viewed by Molly, already suffering with a depressive illness and vulnerable due to her age, affected her mental health in a negative way and contributed to her death in a more than minimal way.

Mr Andrew Walker, HM Coroner and senior coroner for the area of Northern District of Greater London

There is no doubt that more can be done by global players in the industry, not least through the positive use of AI to remove and control content in real-time. But the threat of AI is real and can have devastating consequences.

The Molly Rose Foundation

Following the tragic loss of Molly in 2017, the Russell family and their friends set-up a charitable foundation in Molly's memory. The aim of the Molly Rose Foundation is suicide prevention, targeted towards young people under the age of 25.

You can find out more about the work of the Molly Rose Foundation, including access to resources and how to donate to their work at their website - https://mollyrosefoundation.org

The Opportunities

The positive features of AI systems are particularly useful in pattern recognition, feature detection, optimisation, and forecasting - so there are huge opportunities and benefits in many areas.

ACTIVITY	EXAMPLE APPLICATIONS
E-Commerce	• Personalised shopping experience • Virtual assistants • Fraud detection and prevention • Accessibility
Operations	• Decision support • Task automation • Performance forecasting • Predictive asset maintenance • Threat detection and prevention • Anomaly detection • Factory goods movement • Facilities Management and maintenance
Planning	• Scenario modelling • Optimal allocation of resources • Scheduling • Risk assessment
Supply Chain Management	• Supply chain resilience testing • Supplier assessment and selection • Supply chain risk management • Supply chain integration • Inventory management, ordering and logistics
Manufacturing	• Design • Robotics • Inspection • Testing • Scheduling • Materials management and selection
Learning and Development	• Customised learner content • Personalised learning • Automated assessment • Accessibility • Chatbots
Transport	• Customer experience • Autonomous vehicles and drones • Road-trains • Traffic management • Ride sharing • Navigation • Logistics management

Consultancy - The Impact of AI

As a consultancy in the field of AI, it would be unreasonable not to comment the impact of AI on consultancy and professional services.

In 2021, estimates of the global consulting services market varied between £525bn to £675bn. With this size of expenditure, as knowledge and understanding of AI increases and becomes more commonplace, the outsourcing of routine tasks using AI systems by clients will (rightfully) disrupt the consulting industry.

AI systems that 'look back' to explore common problems and provide independent advice, typically the work of consultants, will become readily available to clients. Combined with AI's capability for search and content generation, at some time in the future AI systems will provide better value services than many consultants employed to undertake the same tasks.

In effect, some of today's consultancy activities would be in-sourced by clients who would use their own AI systems, rather than consultants, for certain activities. These might involve:

- Research and analysis
- Professional and technical advisory services
- Report and presentation content creation
- Project management

In today's consultancy industry, services are reliant on people doing the work with the consultancy business model based on charging clients for the 'time and materials' expended. The replacement of consultancy services that can be delivered by client-owned AI is a significant threat to the current level of revenues in the consultancy industry.

This inevitable change in the balance of capability will demand a change to the type of value created by consultants. Consultants will need to add more value using their human intelligence, knowledge and understanding of context - but AI-enabled in-sourcing by clients means they will pay a lower overall cost for the equivalent service outcomes.

With a proliferation of consultancy since the 1980's and creation of an industry in its own right, concerns are already surfacing about the value-add of consultancy services given their business models. Much of this growth in consultancy has been around the location of capability - client vs. consultant - and using AI is one of the means that a client's internal capability can be increased at much low cost.

In this scenario the balance changes between the value-add of external consultancy and the value-add of internal AI systems adopted by the client. But their combination should lead to an increase in collective intelligence.

"The challenges we face today demand ambitious responses, from the climate crisis to population health. We can do this if government, businesses and civil society foster collective intelligence and mutualistic capacity".

Marianna Mazzucato and Rosie Collington – The Big Con (2023)

A Final Thought

In the same way the 18th Century industrial revolution made an irreversible impact on businesses, jobs, society and the economy, AI will do the same. The industrial revolution led to the replacement of a manual agricultural economy and one where products were made by hand - to a manufacturing economy with increased production capacity at lower unit costs.

People moved on mass to urban areas, their jobs changed, and they had to learn new skills.

As AI becomes more ubiquitous, similar scale economic and societal changes will be triggered.

Both low-skilled and medium-skilled jobs will be lost, especially in the service industry. The innovation cycle will be shorter with more products and services coming to market much faster.

Business models will need to adapt.

A deeper understanding of AI will be needed and the purpose of this paper has been to help leaders who are not working in the field of AI take first step on the journey to mass AI-readiness.

Find Out More

If you would like to find out more about how Interimconsult can help your organisation understand and prepare for the impact of AI, please get in touch.

Paul Corcoran
Managing Director
Interimconsult

e: paul@interimconsult.co.uk

w: https://interimconsult.co.uk