Human-Centred Responsible AI in Modern Organisations

Applying Human-Centred Responsible AI in enterprise systems to empower employees and optimise customer experience

AKKODIS

Introduction

As Artificial Intelligence (AI) systems are gaining an increasing foothold and uptake in predictive performance, their adoption is spreading across several domains. This includes high-stake domains where full automation is not currently desirable due to safety, ethical or legal concerns.

There is growing interest in the AI community to extend human decision-making with AI assistance into organisational processes. After the release of ChatGPT, Bard and other large language model (LLM) based chatbots, the technology has drawn immense interest in foundation models globally. There is much debate around responsible AI due to the 'black box' nature of models and their ever-increasing intelligence.

Akkodis helps organisations incorporate the emerging field of Human-Centred AI decision-making into their delivery methods. The process employs an empirical approach to form a strong foundational understanding of how humans interact and work with technology to take actions and make decisions.

At Akkodis, we believe there are some key considerations to keep in mind in the application of Human-Centred Responsible Artificial Intelligence (HCR-AI) into relevant operational use cases. These are:

- Co-creation of AI models to solve today's practical problems will result in greater trust and applicability of human values into the design and development of AI systems.
- HCR-AI based systems should not only benefit people and societies but should also reduce potential harm by mitigating risks, which in turn will ensure adherence to ethical principles.
- A pattern-oriented reference architecture grounded in design thinking to provide responsible 'AI by design' solutions can address challenges in customer experience.
- Simplification of decision-making can deliver positive outcomes in terms of both customer experience and brand loyalty.

This paper provides insight into the foundations of HCR-AI and management considerations around its practical application across organisational use cases.

Contents

1—AI for Humanity

We should not be fearful of the AI future; in fact, we should capitalise on the extraordinary human creative potential formed by a lengthy cultural evolution and help build well-designed super-tools that boost human performance.

Thought & Communication

As humans, we have an innate need to seek answers and explanations whenever there is a void in our understanding of anything. This is because our understanding is meant to serve a set of goals that predict future events, diagnose problems, resolve cognitive disagreements, and rationalise an action. We have this skill inbuilt from approximately 70,000 years ago when humans discovered new ways of thinking and communicating. The scientific term is 'cognitive revolution' which led to the change in the inner wiring of our brains to think and communicate in much more creative ways.

When we interact with computing technologies, it is natural that we look for an appropriate understanding of how the system works, this is referred to as a user's 'mental model'. This mental model is the very foundation for us to correctly anticipate a system's behaviour and our attempt to interact effectively therewith. What we see and experience with the system, shapes our deep understanding of the structure of reasoning.

In the past few years there has been an increasing adoption of AI technologies, especially machine learning (ML) models, which have begun to be embedded in systems. Due to the sometimes-uncertain nature of ML systems, it can be difficult to understand system behaviour.

As a result, AI ethics is currently a popular topic, and regulation continues to increase. The negative press around the unintended consequences of poor implementations has also not helped the case for AI.

In comparison to previous technology trends like big data, where the focus was on widespread implementation, AI beckons human scrutiny to ensure ethical use. If these concerns are left unaddressed, it may lead to undesired or unintended outcomes. The current challenges arising from the popular debate are the perceptions of what is considered 'Good' or 'Bad' when operationalising AI and the disagreements that arise from this debate.

As Daniel Pink describes in his book Drive: "The opposite of autonomy is control. And because they sit in opposite poles of the behaviour compass, they point towards different destinations. Control leads to compliance; autonomy leads to engagement", and this drives us to 'Mastery', the very need to get better by being better informed."

We should not be fearful of the AI future; in fact, we should capitalise on the extraordinary human creative potential formed by a lengthy cultural evolution and help build well-designed super-tools that boost human performance. In the current evolutionary phase of AI, it is essential that we maintain appreciation for the richness, diversity, and creativity of humans. By not blurring the boundaries between, and clarifying the distinctions between people and computers, we can increase respect for human responsibility whilst guiding people in appropriate ways to use computer power.

2—Building Blocks of AI

To place Human-Centred Responsible AI (HCR-AI) in context, it's important to understand some of the building blocks in what can be a complex picture.

Large Language Models (LLMs)

The release of Chat-GPT, Bard and some other large language model (LLM) based chatbots has drawn significant attention to foundation models globally. Foundation models like LLM are massive AI models that are pre-trained on vast amounts of data and can be adapted to perform a wide variety of tasks to improve productivity. They can also be designed to process and understand natural language and are widely accepted as foundation models that will serve as the building blocks of most future AI systems. Models such as this can generate human-like text, answer questions and complete tasks that involve understanding a language translation and sentiment analysis.

For these reasons, LLMs are a powerful tool that can be used to develop explainability (X(AI)) systems that incorporate natural language processing (NLP) tasks.

Machine Learning (ML)

A newfound ubiquity has allowed Machine Learning to transition from the bounds of academia into an industry engineering discipline. ML is a subset of AI that uses self-learning algorithms that derive knowledge from datasets to predict outcomes.

To decipher the logic and reasoning behind complex ML algorithms, there are interpretability tools designed to help data scientists and machine learning practitioners better understand how ML models work. There has been limited assessment on the effectiveness of tools in achieving their goals, mainly because this field is relatively new and still evolving. The concept of interpretability, which aims to provide reassurance of accuracy, depends on our epistemic understanding of "Justification."

It is crucial to understand how and why AI algorithms make decisions, especially in high-risk domains like cars, military systems or intensive care units where blindly relying on outputs can be dangerous. Therefore, interpretability is undergoing experiments and scrutiny by the academic research and Science community.

Despite these tools' strength in explaining model behaviours, there have been scenarios where explanations were found to be misleading users into trusting biased or incorrect models, leading to faulty models. Biased models could result from existing prejudices, inappropriate data samples, choice of algorithm and confirmation bias.

Mitigation steps must be overseen by independent external 'AI Governance' review bodies as organisations build consequential HCAI systems and venture into unknown territories. It is pivotal to select appropriate training data that is large and diverse enough to counteract common types of biases. Also, the use of methods such as subpopulation analysis for monitoring model performance over time can help to identify if models discriminate against certain sections of the population and prove beneficial in monitoring the health of algorithms.

Machine learning models have found application in critical domains ranging from criminal justice to healthcare systems. Because they have the potential to affect people's lives, it is crucial to understand the behaviours of these models,

so we can be sure they achieve the intended outcome, as well as build a capability to debug and refine models.

Explainability & Interpretability - why do we need it?

The definition of explainability and the related interpretability for machine learning (ML) models is evolving and the subject of ongoing discussion. AI(X) is most prevalent with probabilistic reasoning systems (example: Neural networks), which makes it increasingly hard to understand how results are generated based on specific inputs although, expandability breaks-down the output logically with reasoning.

The most common type of model transparency practiced in the industry are the post-hoc explanations derived from 'black box' models. These often consider a trained model's inputs and outputs to identify patterns in how the model makes decisions. In terms of HCR-AI, explainability plays an important role in the promotion of transparency so relevant stakeholders can understand the decision-making process of the models and potentially improve trust and mitigate downstream harms.

3—Management Considerations Around the Fully-fledged, Practical Application of HCR-AI

Digital supremacy in Customer Experience (CX) has gone beyond the boundaries of traditional customer name personalisation in emails to fully immersive data sets for every customer and their persona.

What's in HCR-AI for the executive?

Before executives can look at AI as a solution, they really need to address the question of the wider vision of an AI strategy which is 'Why are they pursuing AI?'. One of the reasons for the success of high performing organisations is that they follow core practices, which have helped unlock business value by linking organisation AI Strategy to business outcomes.

Business outcomes are closely linked to customer outcomes and organisations across all industries are putting personalisation at the centre of their enterprise strategies to ensure an optimal customer experience. Personalisation and seamlessness are the key competitive 'wrappers' around omnichannel customer experiences. For these reasons, early adopter executives have invested in building intelligent experience engines on the foundations of rich customer data. These engines are providing high-quality customer experiences using AI technologies. This applicability of AI to workflows has led to an increase in employee productivity. According to recent research[i], employees of companies that have adopted AI technology viewed it positively and agreed that it aided in comprehensibility, transparency, reliability, usefulness and operability in their workflows.

To empower their workforce in this way and deliver cutting edge personalised customer experience, leadership need to ensure their organisation's data architecture and strategy is sound to accommodate innovation in AI. Adopting an enterprise data strategy is pivotal to achieving critical business outcomes—underpinned by strong enterprise governance and data classification for HCAI systems. The gap between principles and practice is an important theme and can be addressed by sound engineering practices, trustworthy certification by independent oversight and regulation by government to advance public interest.

A smart enterprise data architecture fuels an organisation's digital and artificial intelligence transformation. Leaders can break the gridlock of traditional architecture design and process complexity by reimagining modernisation efforts. Organisations can choose valid AI use cases and utilise existing reference data architecture that provides agility, flexibility, and innovation. Instead of viewing modern data architecture from a 'waterfall' project lens, organisations can realise results faster by deploying a minimum viable product (MVP) that delivers the specific data components required for a desired use case and continuously improve on the previous versions.

From there, organisations should focus on operational use cases that can be enhanced via AI and engaging non-technical employees in iteratively co-creating AI applications. With the application of HCR-AI practices, leaders can ensure better decision-making while humans remain the final decision-makers in high-risk workflows.

Customer-centric thinking has shifted how organisation's view CX. Beyond simply 'making the sale', customer centricity is a critical success factor for every internal and external touchpoint. Services must cater for a consistent digital and organic experience across the omnichannel i.e., face to face interactions, contact centres, mobile apps and desktop methods, that are seamless and authentic whilst innovative and simplistic. Building HCR-AI into an organisation's wider CX approach in a mature fashion will ensure strong and consistent value realization into the future. Organisations that create enduring impressions on both customers and employees are those that prioritise customer-centric approaches at every stage of AI decision-making. This involves tracking customers across various channels, utilising data insights to comprehend issues and touchpoints, and leveraging HCR-AI technology to craft services that effectively meet unaddressed customer needs.

Lack of Fact-Checking in Generative AI Solutions

A system like Chat-GPT can be applied within any organisation, but it's still a question of 'rubbish in, rubbish out'.

Chat-GPT (a generative pre-trained transformer) is the fastest growing software application in history, it took all of five days for it to reach 1 million users and 2 months to reach 100 million users[ii]. The hype surrounding this growth is leading to organisations rushing to jump on board. However, its limitations must be understood.

Generative AI tools like Chat-GPT in their most basic form under the hood are homogenising free content, text available on the internet, summarising the vast volume of information and presenting a generated output as its creation. If a human was to do the same, it may in some circles amount to plagiarism.

The underlying issue here is that there is no fact-check mechanism employed in this process of information collation and generation. It is a software program that does what it needs to do based on the available pool of information on the free web, irrespective of data quality (i.e., fact or fiction). In other words, without management and oversight, it is much like the Wild West, and Chat-GPT is the 'movie of the month' of the AI world. However, whilst it is going through an exceptional period, it may be inadvertently extracting sensitive or incorrect data and injecting this into its outputs.

We are at the cusp of something transformative. However, rather than being drowned by the initial lustre of this technology, organisations should build their Generative AI approach upon the foundation of strong Human-Centred principles of ethics, defining frameworks and fact-checks to confidently rely on the generated outputs.

HCR-AI: Humans in Charge

Ethical considerations demand that a human oversees AI, not the other way around.

Ethics is a critical component of model design, and it involves the socio-cultural factors spanning a range of AI values. Aligned with responsible AI is the need to make it more human-centric. The intent of Human-Centred Responsible Artificial Intelligence (HCR-AI) is to bring people and human values more closely into the design and development of AI systems in such a way as to benefit society and mitigate harm.

Understanding these socio-technical and environmental factors can help surface the 'why' and 'how' of AI system design to be Human-Centred . Researchers[iii] have found that people may put more capability trust in an AI decision-maker, whereas they may put more moral trust in a human expert (i.e., being able to be morally trustworthy and make decisions that are aligned with moral values as a 'human').

People may not judge humans and machines equally; they may judge humans by their intentions and machines by the outcomes. This does not mean every single decision by a human should be validated; instead, an overriding decision model with specific stages must be incorporated, and explainability built in for scrutiny.

By implementing an HCR-AI approach, leadership ensures that their organisation's programs provide default human oversight and authority, even if AI systems take autonomous decisions on their behalf.

Computer autonomy is a widely used term to describe an independently functioning machine, not directly under human control. In a medium to high-risk domain, if left, unmonitored autonomous decisions may result in failures in usability, reliability, safety, fairness, and other moral crises of AI. The 2019 crash of the Boeing 737 MAX was caused by autonomous manoeuvring characteristics augmentation (MCAS) systems, which took over some aircraft controls without notifying the pilots.

The enthusiasm for fully autonomous systems must remain high and is a valuable research goal. It highlights that humans and machines are embedded in complex organisational and social ecosystems, making interdependence an important goal to maintain and build around.

AI – Governance & Risk Management

There has always been a governance imperative that could impact customer experience, but now Generative AI is introducing new governance considerations that need to be observed from a CX perspective.

As AI techniques proliferate within organisations, it becomes critical to incorporate governance and oversight as the leading attributes in critical areas such as information privacy law to ensure appropriate structures are put in place to mitigate risk.

From a customer experience lens, in addition to digital strategy and culture, an organisation's technology and operations are two key areas of change for a successful digital transformation. Good leadership and governance ensure a well-defined system for decision-making, compliance, and overall stewardship over the organisation's CX operations.

Generative AI may plug into several CX use cases but comes with a variety of risks that need mitigation like legal, regulatory, and specifically 'newer' risks such as that of hallucinations or biased information, where some current Generative AI apps are known to provide factually inaccurate information as output.

An added layer of governance over AI tools is required to ensure, among other attributes, ethics and security. Although ethical principles have been formulated, such as those outlined in the IEEE's 'Ethically Aligned Design' standards, scholars have identified gaps between principles and practice. Based on the principles of a four-layer governance structure[iv] for HCR-AI systems, including: 1. Reliable systems based on sound software engineering practices; 2. Safety culture through proven business management strategies; 3. Trustworthy certification by independent oversight'; 4. Regulation by government agencies can help organisations bridge the gap while building required processes for software engineers, manager, external reviewers and government agencies.

The implementation of external review boards could lead to 'trustworthy' certification and the independent audit of products and services, which would provide a deeper understanding of obligations and thus assist in proliferating successful practices. These practices can support a trusted infrastructure to investigate failures, implement continuous improvement and increase public confidence. Organisations such as the AAAI[v], ACM[vi] and IEEE[vii], have ethical codes of conduct for their members and can hold them to account, however the allocation of liability is a complex legal issue in most cases.

Governments have a regulatory responsibility for safeguarding public interest through ensuring the reliability, safety and trustworthiness of Human-Centred AI systems. Government involvement in AI application is a controversial topic; however, there have been success stories, specifically in transport safety and General Data Protection Regulation (GDPR) in Europe, where collaboration has brought about innovation in explainable AI.

Government policymakers need to be informed about how HCR-AI technologies work and how business decisions affect public interest. Parliamentary legislation governs industry practice, but government agency staffers must make the difficult decisions to help shape the laws. This is where oversight boards such as professional societies or NGOs can advise government officials. Governance structures are a starting point, but newer approaches will be needed as technologies advance or as market forces and customer experience reshape products and services to catapult these to success.

While the risk landscape for AI is fast-changing, the key lies in 'right-sizing' governance as a starting point. Organisations do not need to sacrifice or slow down AI adoption, instead they can utilise existing governance constructs with new AI specific ones like ethics review, bias testing, safety, and explanability.

CONCLUSION

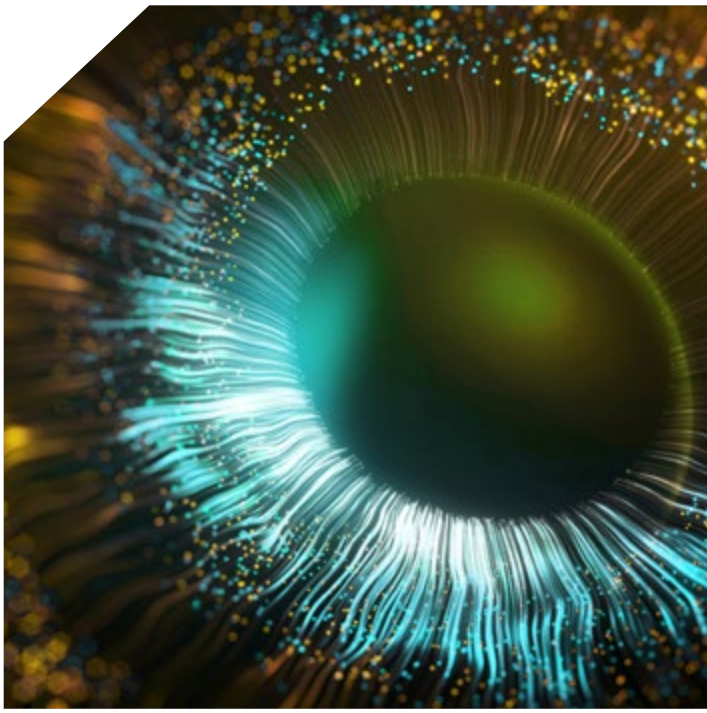
Human-Centred Responsible Artificial Intelligence (HCR-AI) aims to bring people and their values into the design and development of AI systems to benefit societies, as well as preventing potential harm.

Ethics in AI involves socio-cultural and technological factors however different stakeholders have varied perceptions and hence prioritise values accordingly. The emphasis must remain on preserving human autonomy and impact from AI systems application on society and individuals.

Organisational leaders must be cognizant of the strategic implications should they delay developing an AI strategy. Where approached with a mature mindset, the application of AI, including Generative AI techniques can expedite continuous improvement of customer journeys, employee augmentation and strategic differentiation. However, it must keep at its core ethical use, human-centricity, and governance to preserve humanity.

Organisations need to assess their data maturity and AI strategies with ethics being the central load-bearing pillar supporting the future. Because of the ambiguous nature of ethics and the multiple views on a 'dilemma', collaborative dialogue between stakeholders is important when designing AI models. Human-Centred AI application is an approach business leaders and executives can adopt to make sure that the people who are part of the discussion represent diverse professional and cultural backgrounds so that co-creation leads to increased trust, transparency, and promotion of human values.

Although organisations need to act now, it is important to temper the enthusiasm by focusing on building systems that have foundations in design thinking to solve the real operational problems of today. HCR-AI is evolving rapidly as should our embrace of suitable Human-Centred approaches that will amplify our abilities and empower the workforce in truly remarkable ways. If well-managed, HCR-AI will enable us to see, think, create, and act with extraordinary clarity.



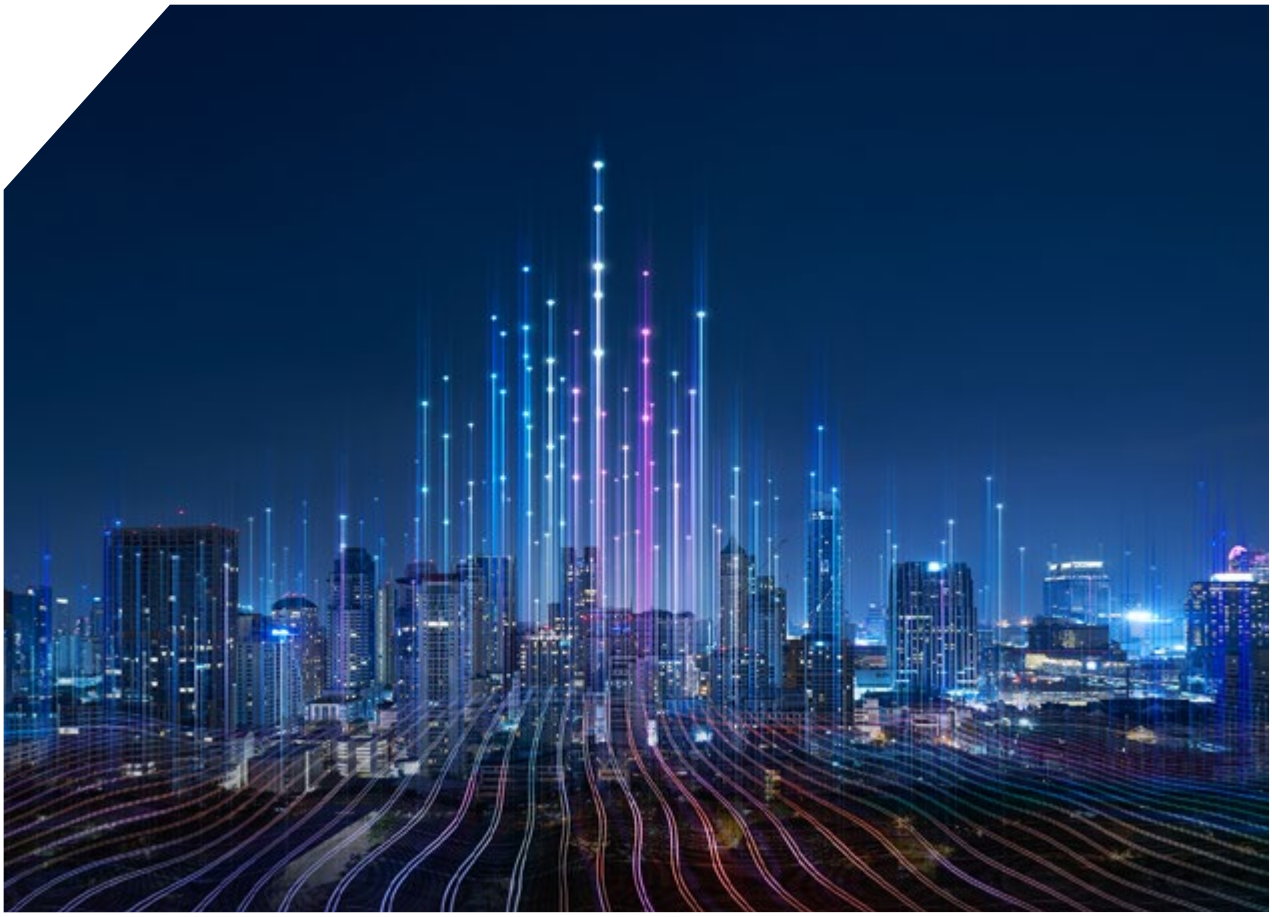
About Akkodis

Akkodis is a global digital engineering company and Smart Industry leader.

We enable clients to advance in their digital transformation with Consulting, Solutions, Talent, and Academy services. Headquartered in Switzerland and part of the Adecco Group, Akkodis is a trusted tech partner to the world's industries.

We co-create and pioneer solutions that help to solve major challenges, from accelerating the clean energy transition and green mobility, to improving user and patient centricity.

Empowered by a culture of inclusion and diversity, our 50,000 tech experts across 30 countries combine best-in-class technologies and cross industry knowledge to drive purposeful innovation for a more sustainable tomorrow. We are passionate about Engineering a Smarter Future Together.



Find out more

If you would like to learn more about how Akkodis can help you apply Human-Centred Responsible AI in enterprise systems to empower employees and optimise customer experience, please contact our team.

Jai Yadav, Management Consultant
Akkodis Advisory
E jai.yadav@akkodisgroup.com

Daniel Roberts, Consulting Director
Head of Akkodis Advisory
E daniel.roberts@akkodis.com
M +61 405 496 175

i Weitz, K., Dang, C. T., & André, E. (2022). Do We Need Explainable AI in Companies? Investigation of Challenges, Expectations, and Chances from Employees' Perspective. arXiv preprint arXiv:2210.03527.

ii Shewale, R. (2023, November 21). CHATGPT statistics: Detailed insights on users (2023). demandsage. https://www.demandsage.com/chatgpt-statistics/

iii Mohammad Tahaei, Marios Constantinides, Daniele Quercia, Sean Kennedy, Michael Muller, Simone Stumpf, Q. Vera Liao, Ricardo Baeza-Yates, Lora Aroyo, Jess Holbrook, Ewa Luger, Michael Madaio, Ilana Golbin Blumenfeld, Maria De-Arteaga, Jessica Vitak, and Alexandra Olteanu. 2023. Human Centred Responsible Artificial Intelligence: Current & Future Trends. In Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23), April 23–28, 2023, Hamburg, Germany. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3544549.3583178

iv Oxford, 2022., 'Part front matter for Part 4 Governance Structures', Human-Centred AI, https://doi.org/10.1093/oso/9780192845290.001.0001.

v AAAI Association for the Advancement of Artificial Intelligence. AAAI. (2023, November 22). https://aaai.org/

vi Association for Computing Machinery. (n.d.). https://www.acm.org/

vii The world's largest technical professional organization dedicated to advancing technology for the benefit of humanity. IEEE. (n.d.). https://www.ieee.org/

viii One example is the collaboration between the U.S. Department of Transportation and Volpe National Transportation Systems Centre to develop an AI software application for automating the detection of grade crossing violations and trespass activities from static camera video feeds. The Grade Crossing Trespass Detection (GTCD) software application outputs predicted grade crossing violations and right-of-way (ROW) trespassing as tabular data along with annotated video files of trespass events. A deep learning-based computer vision tool was developed for automating detection of grade crossing violations and railroad ROW trespass activities. The project successfully demonstrated that a computer vision model can be developed and generalized across multiple locations.

Engineering a Smarter Future Together

Leveraging the power of connected data to accelerate innovation and digital transformation.

AKKODiS

Perth | Sydney | Melbourne | Canberra | Brisbane | Adelaide
T: (08) 9324 8400
E: salesoperations@akkodis.com

akkodis.com