# KNOWRISK CONSORTIUM

## Supply chains: know your risk

December
2021

CATAPULT Digital

# Index

Select one of the links opposite to jump to each section of the report. Throughout the document, there are navigation tabs at the top of the page, so you'll be able to navigate through sections in the same way.

# KnowRisk:
# REPORT – INTRO

# KnowRisk Report

## Sector challenges and insights into mitigating risk for supply chains using advanced technologies.

**Supply chains are in critical need of a redesign - a problem that has been felt acutely during the pandemic. Supply chains often leave a black hole for commercial property insurers, while the companies within these chains are never fully aware of their risks.**

Digital Catapult is part of a consortium that aims to reduce the risk and impact of supply chain disruption through the KnowRisk project. The KnowRisk project utilises artificial intelligence (AI); distributed ledger technologies (DLT); and geospatial intelligence (GEOINT) to collect, analyse and verify risk insights, acting as a proof of concept for future innovation.

Digital Catapult's technical contribution to the KnowRisk project falls into two streams of work:

- The development of an open-source federated learning library for use by the consortium for privacy-preserving distributed machine learning.
- The application of the federated learning library and of a bespoke machine learning (ML) model, to extract risks and mitigations from insurance risk reports.

Digital Catapult has conducted a series of reports on behalf of the KnowRisk consortium to help understand the issues, themes and drivers associated with mitigating risks to supply chains.

This report includes findings and recommendations based on programmes and activities completed by Digital Catapult as part of the wider KnowRisk project, including:

**Revising the Digital Catapult Ethics Framework:** Adapting this framework to reflect a consortium project - being a multi-technology, multi-stakeholder environment. This work enabled the consortium to identify and appreciate the risks that could result from the KnowRisk platform.

**Developing a collaborative ethics roadmap:** Created by independent ethics consultants for the consortium to use, regular discussion and feedback from the consortium transformed this roadmap into a collaborative document.

**Evaluating and adapting applied AI ethics tools:** Selecting Model score cards for federated model reporting and record on negative impact (RONI) to operationalise ethics principles within Digital Catapult's technical contribution to the KnowRisk project.

**Leading policy engagement sessions:** These sessions involved government departments, regulatory bodies and businesses from the insurance, audit, construction and food and drink sectors, resulting in several findings that will help the KnowRisk project provide a beneficial solution for industry.

# Introduction to KnowRisk

In today's globalised economy, supply chains are highly complex. A single problem in one part of the network can impact a multitude of businesses, resulting in risks that are highly fluid and dynamic.

We currently have little visibility into these risks. A business will have one view of its own risk, while auditors and lawyers will have yet another. Meanwhile, insurers will have a detailed view of 5% of commercial sites but are left with statistical models for the other 95%. Currently, these fragments do not come together to create a more holistic view of risk, while an overview of the flow of goods and services through the end-to-end supply chain does not exist. Many supply chain businesses cannot access insurance and collectively suffer $500bn in uninsured losses per year, with many facing unnecessary closure.

The KnowRisk consortium has been created to solve these challenges in the supply chain. Using the latest technologies, it aims to bring together a business' own internal data alongside accounting, insurance and legal (AIL) data, which is augmented with geospatial data, IoT data and over 300 third party data sources to create a 360-degree view of risk.

**Today, businesses deal with a problem reactively. Using KnowRisk's real-time data, companies can collaboratively manage risks and proactively reduce their frequency and impact.** By creating visibility into an individual company's risk and the risk right across the supply chain, KnowRisk can help organisations avoid problems, ensuring they have the right insurance in place for when things do go wrong.

The KnowRisk project has combined artificial intelligence (AI), distributed ledger technologies (DLT) and geospatial intelligence (GEOINT) to serve this critical pain point for insurers across the supply chain and move towards adaptive and robust supply chains. The partner companies involved in the KnowRisk project are **SweetBridge, Engine B, Cystellar, Digital Catapult, Industria Technology and Intelligent AI**,with Sweetbridge being the leading partner.

For more information visit:
www.knowriskconsortium.com

# Summary of reports

Through the KnowRisk project, Digital Catapult is part of a consortium that aims to reduce the risk and impact of supply chain disruption.

Digital Catapult has conducted this report on behalf of the KnowRisk consortium to help understand the issues, themes and drivers associated with mitigating risks to supply chains.

This document is a combination of five reports published to inform the KnowRisk project, the larger industry, and policy development.

## **Know**Risk: Ethics report

**Implementing ethics in practice: ethics for supply chain risk identification, commercial property insurance and the advanced technology that underpins it**

With any new technology, there are always ethical issues to consider. As the KnowRisk pilot combines various technologies, includes multiple players and has the potential to impact several companies financially, ethical concerns must be identified and addressed head on.

**Read this report**

## **Know**Risk: Ethics tools report

**Operationalising ethics principles using applied ethics tools**

This report focuses on applying AI ethics tools to operationalise ethics principles within Digital Catapult's technical contribution to the KnowRisk project, describing how the chosen tools were selected, adapted, used and evaluated.

**Read this report**

## **Know**Risk: Construction report

**Insights from policy engagement sessions**

Although several industry areas of the UK have suffered from stagnating productivity levels in the past decade, construction is perhaps one of the hardest hit. In order for the KnowRisk project to offer a beneficial solution to supply chains, it is important to examine risks within the construction supply chain and identify barriers to the introduction of new technologies.

**Read this report**

# Summary of reports

Through the KnowRisk project, Digital Catapult is part of a consortium that aims to reduce the risk and impact of supply chain disruption.

Digital Catapult has conducted this report on behalf of the KnowRisk consortium to help understand the issues, themes and drivers associated with mitigating risks to supply chains.

This document is a combination of five reports published to inform the KnowRisk project, the larger industry, and policy development.

KnowRisk

## Food and drink report

**Insights from policy engagement sessions**

Between 2020 and 2021, the UK food and drink industry encountered major disruptions - from COVID-19 to the March 2021 Suez Canal blockage - significantly increasing the need to assess supply chains. Digital Catapult led policy engagement sessions on this topic, to help the KnowRisk project to provide a beneficial solution for the food and drink industry.

**Read this report**

KnowRisk

## Federated learning as a service

**Addressing some of the key challenges organisations face when adopting a federated learning approach.**

A federated learning approach is essential when one or more data owners need to adopt machine learning solutions that are trained on and run using distributed confidential data. This report outlines a federated learning as a service (FLaaS) offering, developed to address some of the key challenges organisations face when adopting a federated learning approach to training a machine learning model.

**Read this report**

# KnowRisk: ETHICS REPORT

# KnowRisk: Ethics report

## Implementing ethics in practice: ethics for supply chain risk identification, commercial property insurance and the advanced technology that underpins it.

This paper and the KnowRisk consortium do not purport to have all the answers to these complex questions. Instead, this aspect of the project explores these issues in real life - experimented with and applied to the pilot that is KnowRisk.

**The KnowRisk project aims to provide insurers and companies with an enhanced snapshot and understanding of risk.** A vast opportunity exists to enhance supply chains using a combination of machine learning, distributed ledger technologies and geospatial data, which would be greatly beneficial to both business and society. However, as with all great opportunities and benefits, there are also significant risks, along with the complexities of balancing tradeoffs in design, development and deployment.

With any new technology, there are always ethical issues to consider. As the KnowRisk pilot combines various technologies, includes many players and has the potential to impact several companies financially, ethical concerns must be identified and addressed head on.

In the case of the KnowRisk consortium, which uses all the above technologies, critical ethical questions have included:

- How can data be used responsibly, to create and evaluate supply chain risk and commercial property insurance models?
- What unintended consequences could occur when using self-sovereign identity within distributed ledger technologies and blockchain?
- Are there strong alternatives to traditional business models that will ensure ethics is embedded throughout a product's lifecycle?
- Could transparency in supply chains negatively impact small and medium-sized enterprises and how might this be mitigated?
- How can consortia be governed to ensure participants engage and consider ethical risks?

# Introduction to ethics in the context of KnowRisk

Responsible innovation and applied technology ethics are not just about having the right answers, but focus on embedding repeatable and auditable processes to ensure that risks have been identified, given adequate thought and there are plans in place to mitigate against them. This procedural regularity[1] identifies success for the KnowRisk consortium.

This report will cover:

- the identification of ethical challenges for the KnowRisk consortium and the engagement and activities initiated to address them
- the ethical problems dissected for this piece of work and the resulting decisions made by the consortium
- feedback on the process involving ethics as a service, and how to improve applied ethics in future applications and deployments of these technologies in this critical area for the UK and global economies

With the resulting documentation of this work, the team hopes that other companies and groups of companies working together with shared goals across a variety of supply chains, can learn from and use this as a reference for their own successes.

[1] See page 9, https://arxiv.org/pdf/2102.09364.pdf for more in-depth information on the concept of procedural regularity.

# The current challenges within technology ethics

## The KnowRisk consortium's practical approach to ethics

This section outlines the current challenges within the field of technology, ethics and responsible adoption, highlighting the importance and need for the approach undertaken by the KnowRisk consortium. This section also details Digital Catapult's previous work in applied ethics and how a modified approach has been undertaken to drive impact for the KnowRisk consortium.

### Making applied ethics impactful

Within the advanced digital technology space, technologies such as machine learning and blockchain are rapidly advancing, while any legislation or regulations governing the use of this technology lag behind.

Given how fast technology progresses, it would be almost impossible to constantly update regulations and cover all grey areas of technology development and application. In the absence of hard governance mechanisms,[2] such as explicitly clear legal rules, there has been a surge in soft governance mechanisms, such as codes, guides and principles to encourage those working within the technology space to consider application parameters.

Typically, these guides and frameworks share values in common and often refer to concepts such as justice, beneficence or autonomy. Still, it is difficult to define what these concepts mean in practical terms and there may be differences in their perception or application depending on cultural or geographical variance, with priorities shifting over time.

These difficulties may leave practitioners with many questions on how to implement these values and institutionalise ethics in practice, into their products and services. In addition, many AI ethics or technology ethics tools are often used or interpreted as one off events, or worse, used as ethics washing, to exaggerate or window dress a company's interest in making ethical decisions. This misuse limits the impact these ethical tools can have on mitigating against the technologies' potential harms and risks.

In response to applied AI ethics, the approach undertaken by the KnowRisk consortium is one of procedural regularity.'[3] Procedural regularity does not intend to provide answers to very complex questions - defining justice, for example - but aims to create repeatable and auditable processes, which engrain responsibility, deliberateness and conscientiousness into product, culture and business models.

With a growing backdrop of distrust towards technology organisations,[4] companies may sometimes use the defence that they were unaware of the impact their choices would have on the technology being developed.[5] Whilst it is impossible to be omniscient, it is hoped that practical ethics will enable practitioners to truly consider, from the outset, the implications of their technology and product development to mitigate against risk. Importantly, Digital Catapult hopes to create positive case studies around the value and commercial benefits that applied ethics can bring to technology.

[2] Consider policies leaving a lot of 'grey areas' for technology usage, for example how GDPR is insufficient to protect individuals' privacy in light of inferences that machine learning models can make. See: https://www.law.ox.ac.uk/business-law-blog/blog/2018/10/right-reasonable-inferences-re-thinking-data-protection-law-age-big

[3] https://arxiv.org/pdf/2102.09364.pdf

[4] https://doteveryone.org.uk/wp-content/uploads/2020/05/PPT-2020_Soft-Copy.pdf

[5] See CNN transcript around Facebook and Cambridge Analytica scandal: http://transcripts.cnn.com/TRANSCRIPTS/1803/30/qmb.01.html

# Digital Catapult’s previous work in practical ethics

## Ethics as a service

In previous work, Digital Catapult experimented with and tested a practical methodology to apply ethics to AI. In particular, the practical work focused on implementing and embedding ethics into early-stage machine learning startups. Digital Catapult developed a framework with their Ethics Committee - an independent group of experts in this field, chaired by Prof Luciano Floridi at the University of Oxford. The framework translates high-level principles into practical questions that illustrate how they are relevant to business, people and technology decisions.[6] The framework comprises seven core principles and each of its core principles has an associated set of questions to facilitate a reflective, consultative and deeply practical approach. The AI Ethics Framework is used to support conscientious decision making and promote responsible, questioning and thoughtful startup cultures.

## Applying this work to a consortium project: KnowRisk

Given this previous experimentation of applying ethics to early-stage machine learning startups, the KnowRisk consortium was keen to apply the same approach to the development of their platform. This exercise in application of ethics to the KnowRisk platform was far more challenging and complex than previously demonstrated with just a single startup. Given that KnowRisk is a multi-technology, multi-stakeholder environment, with a number of companies working on different technologies for the same mutual goal, it was necessary to amend the methodology and process to meet the needs of the consortium.

As ethics advisors, Professor Burkhard Schafer and Dr Laura James drove the ethics journey for the consortium. They designed and delivered a number of ethics workshops with the consortium as a whole, as well as for individual company members, to discuss the risks on a macroscale and other company specific concerns. These risks will be outlined in greater detail in the following sections: **Core activities and areas of ethical concern for the project** and **The Ethics Roadmap**

[6] https://assets.ctfassets.net/nubxhjiwc091/xTEqMcYudwQ7GHZWNoBfM/c2a2d55a0ee1694e77634e240eafd-fdf/20200430_DC_143_EthicsPaper__1_.pdf

# Core activities and areas of ethical concern for the project

**Within the KnowRisk project, ethics serves a central utilitarian and commercial function, as well as a practical conduit for harm and risk mitigation.**

Ethics has been embedded into the development and coordination of this project, alongside its objective for sustained, long-term success.
This process encourages all different parties to make thoughtful decisions, examine any unintended consequences and justify their approaches in development.

At its heart, the KnowRisk project has a constructive ideal of responsibility - doing good where possible (as opposed to a constrained ideal, which strictly follows legal rules) and is committed to embedding these values through design. It is hoped that this approach is a tangible method of mitigating short and long-term risk with the avoidance of myopic choices.

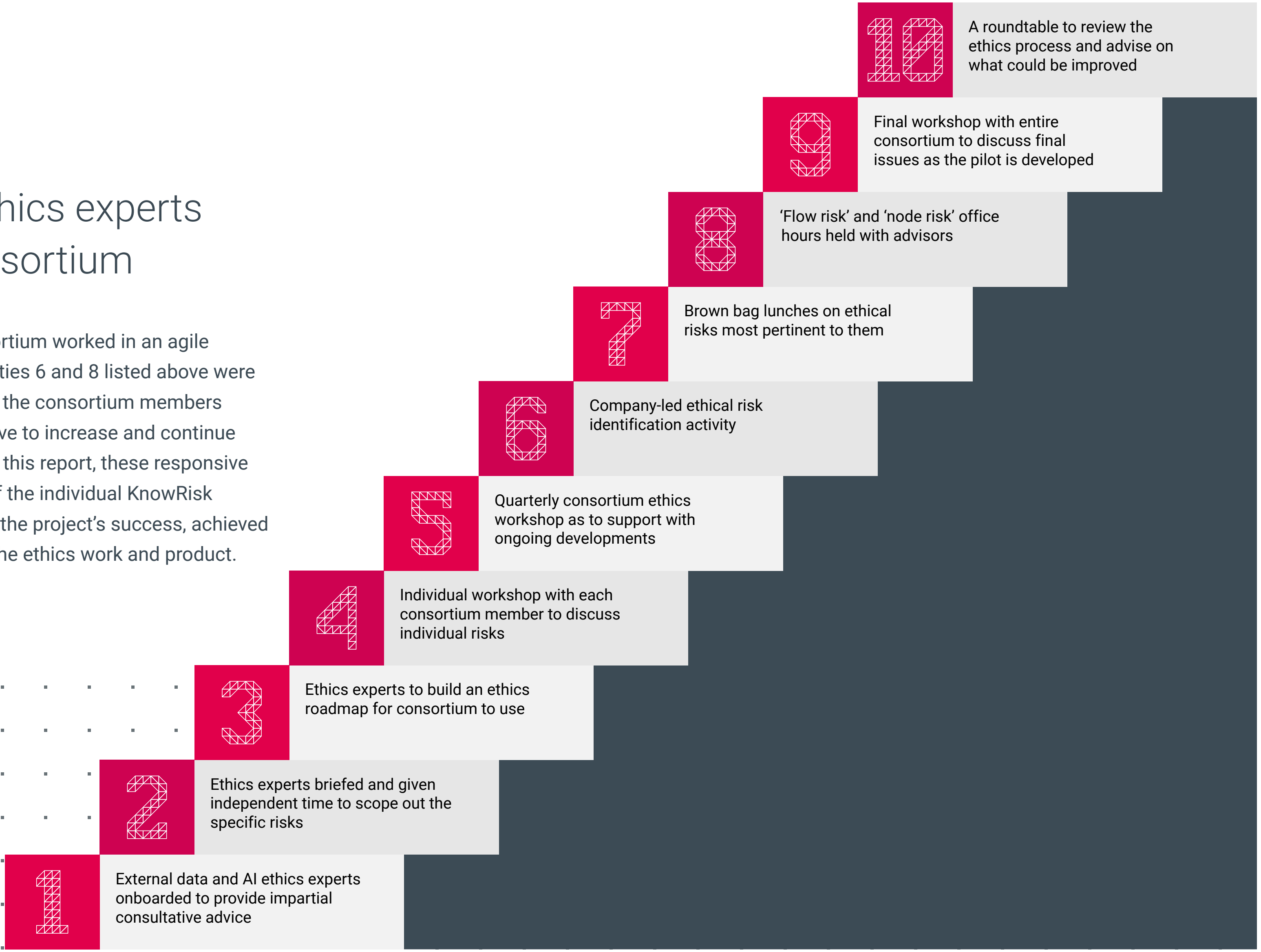
**This section outlines:**

1. The sequence of activities undertaken for ethics as a service
2. The independent ethical advisors and their professional backgrounds
3. Development of a bespoke ethics framework
4. Key areas of ethical concern for the project

# Sequence of activities

## Led by independent ethics experts and the KnowRisk consortium

Throughout the project, the KnowRisk consortium worked in an agile manner, iteratively. As a consequence, activities 6 and 8 listed above were not originally within the plan for delivery, but the consortium members included these activities as a productive move to increase and continue engagement. As discussed in more depth in this report, these responsive tweaks with the enthusiastic participation of the individual KnowRisk consortium members greatly contributed to the project's success, achieved through their attentiveness to the needs of the ethics work and product.



1 External data and AI ethics experts onboarded to provide impartial consultative advice

2 Ethics experts briefed and given independent time to scope out the specific risks

3 Ethics experts to build an ethics roadmap for consortium to use

4 Individual workshop with each consortium member to discuss individual risks

5 Quarterly consortium ethics workshop as to support with ongoing developments

6 Company-led ethical risk identification activity

7 Brown bag lunches on ethical risks most pertinent to them

8 'Flow risk' and 'node risk' office hours held with advisors

9 Final workshop with entire consortium to discuss final issues as the pilot is developed

10 A roundtable to review the ethics process and advise on what could be improved

# Independent data and AI ethics experts consulted

Raising ethical concerns within a company is often met with resistance, sometimes even the persecution of the whistle-blowing individual.

For the partners of the KnowRisk consortium, it was important to engage experts who were impartial and independent of the product development itself. It is not controversial to state that raising ethical concerns within a company is often met with resistance, sometimes even the persecution of the whistle-blowing individual. This approach looks to create a system and dynamic, through which the ethics consultants work as trusted external and independent ethics advisors, with the interests of the collective consortium in mind and it cannot be left to the discretion of any one company to engage or disengage from their advice or recommendations.

It is important to note that these consultants are merely advisory in nature and businesses are not obligated to take their advice. Conversely, experience shows that businesses do prefer to follow a lot of their advice as their recommendations typically improve their product quality. The structure of the engagements, with quarterly check-ins from the advisors, is intended to create a long-standing culture of ethics within the consortium and individual companies.

**Dr Laura James**

*Entrepreneur in Residence at the University of Cambridge*

Holding a PhD in Engineering from the University of Cambridge, Laura James works with emerging technologies in new and growing organisations across sectors. She has been active in the tech responsibility space since 2016, with a focus on effective ways to improve industry practice. Working with businesses and learning about their technologies, challenges and opportunities has always been fascinating to her and she enjoys supporting early stage and growing organisations. Laura is very experienced in enabling startups and scaleups to act responsibly with regards to their users, broader society and other stakeholders, as well as exploring the tradeoffs and choices they face.

**Prof Burkhard Schafer**

*Professor of Computational Legal Theory*
*at the University of Edinburgh*

Burkhard Schafer studied Theory of Science, Logic, Theoretical Linguistics, Philosophy and Law at the Universities of Mainz, Munich, Florence and Lancaster. His main field of interest is the interaction between law, science and computer technology, especially computer linguistics: how can law, understood as a system, communicate with systems external to it – be it the law of other countries (comparative law and its methodology) or science (evidence, proof and trial process)? As a co-founder and co-director of the Joseph Bell Centre for Legal Reasoning and Forensic Statistics, he helps to develop new approaches to assist lawyers in evaluating scientific evidence and develop computer models which embody these techniques. Prof Scharfer has a special interest in the development of computer systems that help law enforcement agencies co-operate more efficiently across jurisdictions, assisting them in the interpretation of the legal environment within which evidence in other jurisdictions is collected. This research is linked to his wider interest in comparative law and its methodology, the idea of a 'Chomsky' turn in comparative law and the project of a computational legal theory.

## Developing a bespoke ethics framework

Digital Catapult's Machine Intelligence Garage acceleration programme has an ethics framework dedicated specifically to early-stage machine learning startups. Using this (experimented and piloted) framework as a foundation, Professor Schafer and Dr James worked collaboratively to update the framework to reflect the needs of an advanced technology consortium. The updated framework also moves away from primarily considering the impact of machine learning, to include additional technologies such as a decentralised and distributed ledger approach to the storage of potentially commercially sensitive information.

The thought process that led to the revision of the framework was to maintain the generality and applicability of the framework, while taking into account any specifics of the KnowRisk ethics work. Unlike ethics consultations for machine learning individual startups, where any harm is likely to be limited to individuals or groups, harm created through the KnowRisk consortium could impact entire market economies or countries.

The amended ethics framework for the consortium is included later in this report.

# Key areas of ethical concern for the project

This section will outline key areas of ethical concern for the KnowRisk project. These were the areas identified at the start of the project to aid the consortium members in thinking about the most pertinent challenges.

These areas include: data and machine learning ethics; supply chain ethics and legislation; power dynamics around transparency; and issues of unrepresented stakeholders. By the end of the project, these issues are explored by the consortium in depth.

See sections:
**The ethics roadmap** and **Interpretation of results and discussion.**

## Data and machine learning ethics

Typically, the law emphasises the areas of focus and concern for companies. The past few years, the ethics conversation has been dominated by questions around personal data and GDPR compliance, where there are severe financial penalties from noncompliance. However, there are still a number of harms that can be derived from using data combined with machine learning, even if this is outside of the parameters within which GDPR operates. Environmental harm is a key aspect of this for example, given the KnowRisk prototype will be used in an insurance context.

If environmentally harmful practices are rewarded with a low economic risk score, it may incentivise or at least facilitate, myopic environmental practices. Previously, we have seen insurance as an industry incentivise all behaviours, some potentially positive and negative - for example, insurance in the past has encouraged people to implement more safety features in their homes with a promise of lower insurance premiums. Therefore, It is important to understand the power of incentivisation that insurance companies can have on individuals.

The ethics advisors on this project have highlighted the importance of understanding data: **data does not always reflect reality; its selection expresses particular concerns, interests and world views. The process of measuring or data collection, in itself, can distort what it is trying to measure.**

This project is particularly privacy conscious: it uses federated learning, where analysis software is run on site where the data is stored, without having to explicitly share data among parties. This enables parties to benefit from an machine learning model that has been trained on a variety of datasets, but still preserves each of the parties' privacy. Whilst there are clear benefits in privacy preservation, its utilisation may not be without its own challenges: how can the efficacy and utility of federated learning be demonstrated and evidenced? Furthermore, how can a consortium using federated learning maintain transparency when the underlying data is only partially visible, as each party can only see their local dataset? It is imperative to ensure that the effectiveness or predictive accuracy of the model is evaluated on an ongoing basis. Providing practical mitigations to these shortcomings of federated learning in a consortium context, through the use of carefully selected applied AI Ethics tools, was the focus of the accompanying **Ethics tools report**.

## Supply chain ethics and legislation

> Leaving the European single market and COVID-19 have highlighted how supply chains are not only essential but incredibly fragile in the face of shocks.

The KnowRisk project commenced, and now continues to develop, in tandem with two major events: leaving the European single market and COVID-19, highlighting how supply chains are not only essential but incredibly fragile in the face of shocks. Currently, social-economic factors mean that supply chains are optimised for efficiency, to maximise the flow of goods, but there is less focus on how to make them resilient.

While this position may be regarded as following the objectives and profit structures of revenue-driven businesses, it starts to become an ethical problem when failures in the supply chain mean that essential services, such as food or medicines, fail to reach countries or cities in the quantities needed. As richer areas or countries might be less impacted than less affluent areas, it is important to create systems, infrastructure and incentives that consider the needs of society as well as company profits.

**It has been also recognised that there are a number of human rights violations or harmful environmental practices within supply chains, often at entry points.**

In 2017, the UK Parliament Committee called for the prosecution of parent companies linked to supply chain abuse[7], where UK companies have neglected human rights in their overseas operations. For KnowRisk, this is an important contextual background. Given the duties of due diligence, regulators might become interested in data made available through the prototype of KnowRisk and, where appropriate, use it against the companies that generated this data.

Consequently, it raises another question around incentives: it is imperative to ensure that companies are encouraged to uncover (and address) human rights violations within their supply chain without being placed at a competitive disadvantage for generating this data.

The questions of how to engage with regulators and how to communicate this with other stakeholders is key, as well as being wary of any new incentives for companies to falsify records, which may impact on the degree in which tools are openly shared.

If KnowRisk's technology becomes used for certification or if third parties use it to certify transactions, this could increase the demand for accuracy, correctness of results and new transparency duties. It could also make it de-facto impossible for participating companies to switch to a different platform, resulting in technology lock-in, which may have a greater impact on companies with fewer financial resources.

## Power dynamics around transparency

While transparency is often referred to as an unalloyed good, access to information does not benefit all stakeholders equally. In the case of supply chains, larger parties might sometimes use transparency to extract profit from smaller ones. This might, for example, happen if players want to undercut other companies in the supply chain or try to exclude them entirely.

There are also risks around economic warfare, particularly in the current geopolitical landscape. As a result of this tension, the KnowRisk consortium would like to clarify that the project does not aim to increase supply chain transparency; rather the aim is to increase accountability without requiring transparency (and the associated problems) in the hope that this will make the project valuable in light of current supply chain issues.

[7] https://www.business-humanrights.org/en/latest-news/uk-parliament-committee-calls-for-prosecution-of-parent-companies-linked-to-supply-chain-abuse/

# The ethics roadmap

The ethics roadmap is a critical tool, developed to lay the groundwork for ethical technology.

The ethics roadmap is a critical part of this ethics workstream and engagement. Many technology ethics frameworks exist, including over 160 frameworks around the ethical use of AI as of 2020.[8] Many of these guides are aspirational or encourage us to consider what we may believe to be right or wrong when it comes to technology and ethics.

While this body of literature has laid the groundwork to bring us closer to what ethical technology might look like, these guides often lack details on how to operationalise these values to create business outcomes. The intention for this bespoke roadmap is for it to be a critical tool to bridge this gap between aspiration and reality.

This ethics roadmap was developed by the independent ethics consultants. It was produced after the first group consortium session and each of the individual company sessions. Throughout each of these discussions, the advisors picked up on prevalent themes and developed actionable recommendations. The consortium was asked for feedback on the roadmap, plus any omissions or amendments, which transformed the map into a collaborative document, outlining what could be accomplished within the timeframe.

## In the first month

Across the consortium

- Continue work to help all consortium members understand and feel involved in the KnowRisk vision and overall aims.
- Consider an informal, storytelling-style remote meeting, where key project leaders can talk through the vision and ambition (with examples and ideas, rather than slides and diagrams), for the whole project team.
- Identify a forum or process by which individual or team concerns, as well as positive suggestions about ethics in the project, can be raised at consortium level and also internally if clear procedures are not in place. This includes ways to assure more exposed members (on short term contracts, with less social capital) that raising concerns can make a positive contribution to the collective effort, presenting it as a task for everyone to identify opportunities for the growth of ethical practices.

[8] See: https://assets.ctfassets.net/nubxhjiwc091/xTEqMcYudwQ7GHZWNoBfM/c2a2d55a0ee1694e77634e240eafdfdf/20200430_DC_143_EthicsPaper__1_.pdf and https://inventory.algorithmwatch.org/

## For November 2020

### Proposed activities:

- Team ethics concerns are raised with Burkhard Schafer (BS) and Laura James (LJ) if it is felt that no route to do so exists within the consortium.
- To build team cohesion, remote informal talks, perhaps organised over coffee or a brown bag lunch, featuring different team members or invited guests, with discussion time, could be scheduled once a or so during the project duration.
- Thought should be given to informal moments where team members can get together and chat - one example would be a remote weekly coffee slot for participants to drop into at their choice.
- Relevant team members across the group should meet to discuss the threat model which impacts data privacy in KnowRisk and document the results.

### Individual partner companies

- All partner companies involved in the KnowRisk project – SweetBridge, Engine B, Cystellar, Digital Catapult and Intelligent AI – should dedicate time (perhaps a couple of hours) as a team working on KnowRisk to review the Ethics Framework and consider how each section impacts the work of the organisation on KnowRisk.

### Across the consortium

- Ethics check-in with Prof Burkhard Schafer and Dr Laura James
- Relevant team members to connect, question and discuss how the ever-evolving nature of machine learning, used in several parts of KnowRisk, might be presented and worked with to users of the project in the future.
- Consider creating materials to explain this to future stakeholders (e.g. an explainer, comic, short video or other form)
- Set up a KnowRisk webpage including a statement on its approach to ethics, contact details and procedures that would allow organisations or individuals, who fear they have been unjustly affected, to ask for remedies.
- To complete at least one stakeholder workshop or conversation - with a risk assessor or commercial insurance buyer, for example. This could either be a workshop with a range of stakeholders or smaller conversations with relevant project team members and one or two stakeholders.

### Engine B

- Deliver an informal remote seminar to the overall consortium project team explaining how Engine B has been set up to balance purpose and profit, while protecting the mission against bad outcomes and actors and so on. This seminar will showcase a different way of shaping an organisation or consortium, which may support thinking for the KnowRisk work after 2020-21.

### Intelligent AI

- Explore issues of debiasing, explainability, transparency and other core ML ethics questions for Intelligent AI's work on KnowRisk, producing a short informal written briefing or presentation for the project.

### Sweetbridge

- Publish an accessible essay, video or other output describing KnowRisk and the Sweetbridge system and model.
- Share some of Sweetbridge's thinking about ethics for its platform and the future of the project, with the whole team and ideally a wider audience online (e.g. an informal webinar talk with a Q&A session).

## For January 2020

### Across the consortium

- Ethics check-in with Prof Burkhard Schafer and Dr Laura James.

### Individual partner companies

- All partner companies involved in the KnowRisk project – SweetBridge, Engine B, Cystellar, Digital Catapult and Intelligent AI – should evaluate their work on KnowRisk and consider how explainable it is to a general audience (a Financial Times reader, for example) and consider writing a short blog post or similar informal article that outlines the completed work to date and how it is appropriate, fair and so on. The Ethical Framework is a useful tool to inform any thinking around this article.

### Digital Catapult

- Explore issues of explainability, transparency, privacy, robustness and other core ML ethics questions for the Catapult's work on KnowRisk and produce a short informal written briefing or presentation for the project.

## By project end

### Across the consortium

- Consider how to clearly articulate the project vision and components in the final reports and other assets (such as any project website or archive) for accessibility to non-expert stakeholders.
- Evaluate the project progress in light of the ideas uncovered in the first ethics workshop focusing on: what the consortium could be proud of; and what would be the worst future headline. Then, consider how these ambitions and fears may, or have been, progressed or alleviated.

### Digital Catapult

- Prof Burkhard Schafer and Dr Laura James input into the final ethics report, with any final conversations and checks within the project completion.

## For future work after this proof of concept project

### Across the consortium

- Review and consider the final ethics report from the 2020-21 project.
- Review and consider the accompanying ethics tools report from the 2020-21 project.
- Review and consider any parts of the KnowRisk ethics roadmap (this document) which were not completed during the 2020-21 project.
- Prioritise ethics and governance questions in the design of any f uture projects, including allocating responsibilities. Establish regular feedback processes to ensure stakeholders are appropriately engaged and on board with the work.

# Interpretation of results and discussion

Responsible innovation represents a long journey; it is impossible to make changes overnight.

This section will outline the outcomes of the KnowRisk ethics project. For clarity, outcomes will be centred around the principles in the ethics framework, developed by Digital Catapult and adapted for KnowRisk:

- Be clear about the benefits of the product or service
- Use data responsibly
- Know and manage the risks
- Be worthy of trust
- Promote diversity, equality and inclusion
- Be open and understanding in communications
- Consider the business model

On the following pages each of these principles, the consortium's efforts and the impact of the interventions will be discussed. In addition to this, there may be ethical challenges which arose out of deeper consultation with the ethics experts in the later stages of the project, of which direct actions are still being undertaken. The thoughts and discussions are documented here as they pose important questions for both ethical issues in technology and supply chains. Responsible innovation represents a long journey; it is impossible to make changes overnight.

This work has looked at ethics from a consortium point of view. This has meant that each partner has had to align with one another on the approach and implementation of ethics. This approach is vastly different to implementing and embedding ethics into individual startups, as companies are typically composed of 2-10 people (usually 2-4 individuals); therefore all the decision makers are often in a room together and have the ability to make quick decisions.

> “Ethics isn't a tick box exercise, it is the life and blood of the organisation. It is a requirement to doing business.”
>
> **Anthony Peake - CEO, Intelligent AI**

The dynamic within startups is also different, mainly as individuals completely understand all stages of development and any potential issues. However, embedding ethics across a consortium is a difficult task: meetings typically only include a few representatives from each organisation, rather than a full contingent and, while these members have a wealth of contextual knowledge, it may be fragmented due to the distributed nature of product development in a consortium. Therefore, a higher level of engagement, bonding and governance is required across the consortium to be successful.

## Be clear about the benefits of the product or service

KnowRisk aims to benefit supply chains and mitigate against disruptions. It provides a first step towards a future with modern, adaptive and robust supply chains for manufacturers, retailers, suppliers as well as insurer, legal and accounting firms.

As the KnowRisk platform has the potential to impact a large number of global supply chains, economies and countries, it needs to be clear who the platform will benefit. For example, the consortium has considered at length, if there could be a disparate effect on different countries and if this would contribute to the gap between richer and less affluent countries.

Given this application is also being used in a global context, it is likely that what might be defined as a benefit to one country, may not necessarily apply universally. For example, while transparency is often seen as an intrinsic benefit, it is far more instrumentally beneficial to a select number of parties: by having complete transparency, retailers could drive down prices of smaller suppliers across the network to uncomfortable or unprofitable levels. The consortium is also acutely aware that they cannot impose their world view into other countries they operate in.

## Know and manage your risks

Identifying and mitigating against risk is critical to the success of technology products and platforms. In Digital Catapult's previous work, it has been identified that having a good grasp of risks can help with securing investment and customers more effectively.[9] Whilst the KnowRisk platform as a single entity is not yet at a stage to start discussions with investors, this has played an integral part in any preparations for the future. Equally, Intelligent AI has successfully raised investor funding during involvement in KnowRisk and they believe their involvement in ethics was useful in the funding process.

From an outsider's perspective, it was interesting to deep dive into how technical personnel consider ethics. When engaging with different companies' technical teams during the initial stages of the consortium, there was sometimes a tendency to defer solely to technical features as providing a solution to ethical problems. For example, if asked, "What would happen if data is leaked for unintended purposes?" the discussion may have focused on specific features that, technically, would make data leaks very difficult to achieve.

Throughout the project, it was notable how these responses greatly evolved and transformed. By the end of the project teams were much more empowered to discuss hypothetical (but possible) ethically-challenging scenarios and how to best manage and deal with them. In particular, this included the type of processes that might be implemented to ensure users can effectively complain about problems using the platform and how to rectify their results. As an example, the Intelligent AI team have implemented bi-weekly show and tell sessions with the development team, whereby they challenge ethical implications and discuss any ethical concerns candidly and openly.

> "Intelligent AI has successfully raised investment during KnowRisk... the ethics work was discussed with investors and I believe that made a difference in raising the funding."
>
> **Anthony Peake - CEO, Intelligent AI**

[9] https://www.digicatapult.org.uk/news-and-insights/publication/unveiling-the-commercial-value-of-the-responsible-use-of-ai

## Use data responsibly

It is important to understand the responsible use of data in a machine learning context, as not only being an ethics consideration, but also fundamental to a high-quality product. The consortium recognised early on that the responsible use of data would also lend itself to the strong predictive accuracy of algorithms, whilst balancing precision. This feature is critical; if users find the platform not working as promised,  KnowRisk would struggle to retain customers, resulting in a high churn rate. Consequently, ethics should be seen as aligning with strong product outcomes.

## Using data responsibly in machine learning aspects of KnowRisk

The machine learning contingent of the consortium have a very rigorous approach to biases in data. Understanding that data bias has no one quick fix - as all data that is collected, has been collected by someone, for a specific purpose, looking at a defined number of variables - is important for robust and repeatable outputs. Data will have a number of historical ills to it - for example, sloppy data labelling may indicate that the area directly outside a police station suffers acutely from higher crime rates than the surrounding areas, but the explanation will be that the data has been labelled incorrectly by the person collecting the data. Equally, there may be biases in higher crime rates outside and near police stations, simply because people are more likely to report a crime if it is easier to do so, which again creates a problem of bias within the dataset.

As a result, the machine learning team, Intelligent AI, has built specific toggles in their data visualisations, to be able to discern whether a result changes due to the input of additional variables. For example, when looking at the impact of value ‘X’ on crime statistics in an area, users of the platform could remove the variable ‘X’ and visualise what other (potentially more pertinent) factors contribute to crime in the same area. This might highlight the wrong attributions of causality to specific problems.

Sweetbridge and the wider consortium also have built ethics into data analytics and visualisation processes - an example being the ability to see raw analytics and ethics-corrected analytics, so clients can see the impact of different features and make informed decisions.

Digital Catapult has also demonstrated the benefits of using a federated learning system to train machine learning models while protecting sensitive data and maintaining a level of transparency using tools like model score cards for model reporting and record on negative impact (RONI), as outlined in the **Ethics tools report**.

## Using data responsibly within blockchain and distributed ledger aspects of KnowRisk

Along with the independent ethics advisors, the consortium discussed at length some of the issues in using self-sovereign identity on blockchain in respect to being responsible with data. There are clear benefits to having autonomy over identity, i.e. not sharing unnecessary excess data, but there are also some ethical issues.

First, as the notion of self-sovereign identity (SSI) is individualistic, it may emphasise the individual too much, rather than encouraging collective action or communal collaboration. There are technical approaches that can mitigate against this individualisation. For example, it can be demonstrated that a person has contributed to a group discussion or engagement, without disclosing the nature of the engagement and then have all individuals who contributed to form a group digital signature on a blockchain, thereby creating a collective vision and the potential for group responsibility. However, this approach may be limited by company cultures.

There are also concerns around biases resulting from who decides to not disclose data. Research indicates that those who are more privacy conscious tend to be more affluent and educated[10] and only these groups tend to exercise their rights within data privacy.

Questions need to be asked: who will benefit most from self-sovereign identities and does it perpetuate a system that continues to protect and provide more security to the most advantaged groups?

In addition, using self-sovereign identity (SSI) in a corporate context may produce other benefits and concerns. Zero-knowledge proof, used by Sweetbridge, enables businesses to ask questions of datasets and get answers without revealing any underlying data and therefore protecting privacy. These answers have proofs, demonstrating to businesses that the answers they receive are correct without being able to see the supporting data. This is a crucial aspect of the system, fully focused around the issue of privacy within supply chains and the protection of commercially sensitive data (protecting the commercial interests of businesses).

As blockchain and distributed ledger technology broadly operate on the premise of decentralisation, there is a bigger question about what happens on a large scale when individuals choose not to share their data. For example, in the case of discrimination and unfair algorithms, how much knowledge is needed to understand that groups are being marginalised as a result? If data is kept secret, we may not have the bigger picture of what groups are being discriminated against and how.

[10] https://www.cisco.com/c/dam/global/en_uk/products/collateral/security/cybersecurity-series-2019-cps.pdf

> “It is likely our product will be (technologically) ahead of regulation. We needed to ensure we built the correct processes to go above and beyond and protect all users.”
>
> **Scott Nelson - CEO, Sweetbridge**

The Sweetbridge team is developing proof-oriented programming, which instead of sharing data enables parties to exchange proof that their data meets requirements and has specific properties. This enables private aggregation, whereby users are able to see aggregated results, but what specific values contributed to the end aggregation remain private. In this context, the ethics advisors spent time with the Sweetbridge team to discuss how developers of this system will know if this is working correctly. Within a legal context, when it comes to litigation, someone would have to demonstrate that the platform hasn’t worked. Consequently, what would be needed to prove a mistake or failure within the system?

This discussion emphasised the distinct notions of evidence and proof; and that evidence should be understood as ‘evidence for a certain thing.’ These sessions evolved the consortium’s approach to the validity of forensic methodology required for system data to be admissible in court.

## Be worthy of trust

There is a paradigm when it comes to trust. Trust experts have defined a combination of behaviours and drivers, such as reliability and competence in behaviour and integrity and empathy as drivers, that make individuals and companies worthy of trust.[11] Therefore, trust is something earned over time through consistent behaviours and the willingness to take action, even if it is deemed as potentially undesirable to business in the short term. Long term, this trust would pay off through the retention of customers and winning new clients or investment.

To consolidate trustworthiness, the consortium took part in a number of activities. For example, each company within the consortium undertook ethics training and workshops for their entire teams. In addition, Intelligent AI is developing personas for user groups and for non-user groups, so that they can test the platform against the impacts of each persona. Intelligent AI is also building machine learning algorithms that are auditable and not black boxes, as well as working to ensure that the platform cannot be used to profile and prevent smaller organisations from gaining favorable insurance rates. Other members of the consortium, CyStellar for

[11] https://medium.com/@rachelbotsman/being-more-trustworthy-the-basics-6354e504917f

example, has implemented ethics into data infrastructure and architectures by providing: automated DevOps checkpoints; automated tests and validation points; accepted and unaccepted datasets; supplier classifications; and specific guidelines to project managers.

To build trust in the platform externally, the KnowRisk project team have engaged with real users throughout the project via the 2Build consortium. This is a group primarily constituted of small and medium-sized enterprises (SMEs) , engaged to ensure that the companies whose risks (through the KnowRisk platform) are being assessed are understood. This thereby limits the risk that key stakeholders are not represented as part of the product development process.

As the KnowRisk consortium works towards commercialisation of the platform by the end of 2021, they are expected to find a third party to collaborate with, such as the social tech trust to enable audits of the agreed ethics standards.

## Promote diversity, equality and inclusion

The Ethics Framework and workshop enabled the KnowRisk consortium to better identify and appreciate the risks that could result from the KnowRisk platform. This has proved important, as for a number of members in the consortium, as doing good was, and remains, a key motivator for them as individuals and companies. Whilst this is undoubtedly a desirable trait to have, companies and individuals may be led to believe that good will and good intentions are sufficient to avoid harm. However, the ethics exercises that the consortium completed as a group encouraged members to think thoroughly about the risks that could derail their goal of being fair and balanced and to ensure that there were processes and governance structures in place that would be robust enough to mitigate against these risks. It underscored the need to establish a proactive and inclusive governance approach that seeks and welcomes input about ethical biases and offers a fair and transparent path to resolution. This is particularly needed in an endeavour such as KnowRisk, where the technology is likely to be ahead of the regulatory bodies' awareness and ability to protect its citizens from possible adverse effects.

On reflection, one particular mechanism that worked well has been the appointment of a project manager for the consortium who took ethics in their stride. In this case, the lead company Sweetbridge established a role specifically to manage the KnowRisk consortium.  While this has been outside of the recommendations in the roadmap and has been formally included in the consortium building agreement, this is a notable way to ensure ethics is being considered effectively. This role ensured that all different stakeholders within the consortium were being listened to, included and encouraged throughout the ethics work. Furthermore, this consortium project manager was able to identify when ethics needed more attention and how to mitigate against the risk of individual companies feeling left behind.

Interestingly, throughout the literature of consortia governance, there is no established best practice on how to delegate and assign project leads. Depending on the objectives of the consortium, different structures might work best. For example, consortia whose main priority is to be as democratic as possible may want to have rotating heads, whereas consortia who hope to make decisions quickly and more effectively may look to have a smaller steering group or an individual managing decisions.[12] Conversely, through the experience of the KnowRisk project, having an established point of contact for the ethics workstream has been useful for effective communication, governance and inclusion of parties.

The importance of this overarching principle is not to be underestimated: strong relationship building and bonding from the outset of the consortium were critical to the success of the consortium. This sense of interdependency and a shared vision across consortium members has been fostered through virtual brown bag lunches, informal meetings held every couple of months where members could discuss the project and updates. The process of having consistent ethics workshops, with independent advisors, was also described as being a fruitful mechanism for bonding, interaction and ensuring members were developing a uniform culture across the consortium. This also proved important for Industria Technology - the blockchain company joined the consortium later than the other partners, yet meaningfully engaged with the ethics workstream as a result of these different consortium engagements. Their enthusiasm and willingness to learn from the ethics experts has been instrumental to them getting up to speed quickly, demonstrating that these mechanisms can help when bringing in new members and ensuring that ethics continue to be instilled across partners.

## Promote diversity, equality and inclusion in consortium governance

New technologies can change existing power hierarchies, so it is important to ensure that the governance within the KnowRisk platform fairly manages power between stakeholders and prevents powerful groups from co-opting the platform to further their own interests.

As a small starting point, the consortium is developing a statement that will be made publicly available on how to engage any stakeholder, to ensure that any one party is not using the data and platform to further their business interests unfairly.

In addition, the consortium has considered the possibility of legally creating a single entity/joint venture agreement, to help align organisations within a collective article of association. However, to quote one of the ethics advisors: “culture eats strategy for breakfast.”

Potentially, the most important aspects to keep ethics alive are mechanisms to ensure it continues to be embedded in the culture. As an example, a few suggestions were:

> 1) creating hard criteria for new board members that would ensure they met some standard of technology conscientiousness.
> 2) running workshops and role plays when setting up future consortia.
> Both of these suggestions could be translated into governance and economic guidance.

[12] https://gh.bmj.com/content/⁴/Suppl_8/e001450

## Consider the business model

By the end of the project, members felt that ethics was proven to be a top priority for business and continues to remain a high priority. Members expressed a desire to keep ethics alive, to be embedded into the product and to not become a victim of ethics washing or other corporate social responsibility exercises, whereby it is merely for window dressing and doesn't do enough to protect against potential harms to users or other groups. To thispurpose, the consortium has considered the importance of legacy planningfor the KnowRisk project, asking how it can be certain that institutional memory is embedded into the consortium, regardless of whether project partners are eventually replaced by other ones.

Practical steps have been taken to identify who the business model impacts. As Intelligent AI's platform is a B2B model, focussing on commercial property insurance, it has less ethical impact on citizens and individuals, but potentially could present an ethical bias against small and medium-sized enterprises and favour large businesses instead. To monitor this, the company is building analytics capabilities into the platform.

**On a broader level, areas of future research and exploration may include research into emerging venture capitalists that are structured in nontraditional methods and do not encourage companies towards an exit.**

It is useful to frame innovation building around the incentives imposed due to financial support in venture capitalists (VCs) and investors, as this has historically created a moral duty on founders to sell their businesses and prioritise extremely high valuations. Second generation business models that may focus on other variables such as revenue share, as opposed to company valuations, may promote the creation of slower but more sustainable and conscientious businesses.

# Feedback

## How to improve ethics as a service

Reflecting on the process, it was felt that the quality of input from both ethics advisors has been extremely high, highly thoughtful and their documents (the roadmap and ethics framework) have become reference documents for the KnowRisk project. Across the board, consortium members felt that the ethics workstream did everything it promised from the outset.

**In terms of constructive feedback, it was felt that the user experience could be improved upon for future consortia or commercialisations of this process as a product or service.**

In particular, the layout of the ethics service, which was predominantly documented in long form, made the barrier to entry quite high, as it proved to be quite difficult to ensure everyone would read, engage with and evaluate the documents. This format was perhaps more academic in nature, giving members the impression that they were being taught ethics. As the KnowRisk consortium had many members who are highly experienced professionals, this approach was potentially not the most effective way to engage them. In addition, the group activities were, initially, all large online workshops. This was done to create the feeling of a single entity as a consortium but due to the larger group size not everyone wanted to contribute openly.

This problem was identified during the course of the project. The KnowRisk consortium manager worked closely with the Digital Catapult team to increase engagement in the ethics activities, developing a risk identification activity,[13] to be held during a brown bag lunch with individual consortium members.

> “The quality of the input from Dr Laura James and Professor Burkhard Schafer was extremely high and the framework and roadmap are highly thoughtful and detailed and have become essential reference documents for KnowRisk.”
>
> **Jason Cresswell - KnowRisk Consortium Manager**

[13] See links at end of report for more information.

This meant that internal teams were holistically engaging with the most relevant questions for their specific technologies in a critical way. These conversations enabled teams to truly understand the ethical issues covered in the framework and beyond. To further this understanding, the Digital Catapult team organised additional office hours sessions with the two independent ethics experts to discuss the issues that came to light during their internal ethics meetings. The sessions were split into two groups: node risk and flow risk, which allowed the teams to dig even deeper into ethical considerations. These sessions took people outside of their comfort zone and enabled them to understand the complexity and competing tradeoffs involved in their decision making.

> Ethics as a service has a long way to go, but every trial and iteration of its experimentation will reap stronger results.

While the high standard of the content offered and produced is not  in question, the way it was presented has potential for improvement to ensure it reaches the highest levels of engagement. It is likely that the most effective way of achieving this also depends on consortium to consortium; potentially, academic consortia might have seen a preference for the initial framing. In any event, the design and user experience of ethics as a service has a long way to go, but every trial and iteration of its experimentation will continue to reap even stronger results.

For future industry-led consortia, the approach of having individual teams run through an ethics identification activity, and then a small number of participants in office hours sessions with the ethics experts to dig deeper and discuss the issues identified,  is recommended as a complement to the larger consortia ethics sessions. It seems both are essential to enable ethics to have an impact at the local level (i.e. within teams) as well as at the consortium level.

# Closing remarks

## For the KnowRisk ethics report

It is evident that cultivating ethics within the KnowRisk consortium has been immensely beneficial, to the culture across teams; quality in product design and development choices; and in its approach to the platform's potential impact on wider society.

Responsible innovation doesn't happen overnight, but it is a worthwhile endeavour. It is especially important to have these conversations in the early stages of projects, as this can set the precedent for how entire projects are governed and technologies developed.

# Links

Revised ethics framework for KnowRisk

Identification of ethical risks worksheet

# KnowRisk: ETHICS TOOLS

# KnowRisk: ETHICS TOOLS

## Operationalising ethics principles through the use of applied ethics tools

**This report focuses on applying AI ethics tools to operationalise ethics principles within Digital Catapult's technical contribution to the KnowRisk project. The report should be read in conjunction with the KnowRisk: Ethics Report, the scope of which is the ethics work across the KnowRisk project and consortium as a whole.**

Digital Catapult's technical contribution to the KnowRisk project falls into two streams of work:[1]

- The development of an open-source federated learning library for use by the consortium for privacy-preserving distributed machine learning.
- The application of the federated learning library and of a bespoke machine learning (ML) model, to extract risks and mitigations from insurance risk reports.

Alongside this technical work and as part of the ethics workstream, Digital Catapult selected, adapted, used and evaluated two applied AI ethics tools with the view to enhancing the transparency and robustness of the federated learning system.

Those tools were:

- **Model score cards for federated model reporting**
- **Record on negative impact (RONI)**

This exploratory work has, to some extent, demonstrated the potential utility of applied AI ethics tools as part of a wider responsible innovation approach. The consortium found the experience of applying these tools alongside the technical contribution from the group to be very useful in terms of expanding the general hands-on experience in applied ethics.

This report details the methodology, experimental design, results and final evaluation of these tools in the context of the KnowRisk project.

[1] https://github.com/digicatapult/dc-federated

## Introduction to project context and the use of AI ethics tools

KnowRisk is a collaborative research and development project. The aim of the project is to develop a platform for organisations to measure, mitigate and price risk for complex modern supply chains.

The KnowRisk project utilises artificial intelligence (AI), distributed ledger technologies (DLT) and geospatial intelligence (GEOINT) to collect, analyse and verify risk insights.

Given the potential opportunities and risks inherent in such a project, involving advanced digital technologies at different levels of maturity, the application of practical ethics has been deemed as essential from an early stage. The Ethics report provides a holistic view of this work while the Ethics Tools report focuses on the identification, adaption, use and evaluation of applied AI Ethics tools as one aspect of the operationalisation of ethics within the KnowRisk project.

Past work completed by Digital Catapult and the Oxford Internet Institute on a typology of AI Ethics Tools[2] provided a starting point for tool selection along with a consideration of the initial ethics deep dive results to prioritise needs. In the view of the consortium, z , detailed in the ethics report, highlighted the following area of specific ethical concern: **maintaining robustness of machine learning processes, while respecting the privacy of sensitive commercial data and the need for some level of transparency with regards to the machine learning models used and the underlying data.**

To address the above concern, and given the distributed nature of both the KnowRisk consortium and the proposed network of users (insurance companies, small and large businesses connected by supply chains and technology providers), a federated learning (FL) approach was adopted.

[2] https://arxiv.org/abs/1905.06876

**Federated learning can be defined as “a machine learning technique that trains an algorithm across multiple decentralised edge devices or servers holding local data samples, without exchanging them.”[3]**

As part of the KnowRisk project, Digital Catapult developed an open source library for federated learning that has been designed for industrial cross-silo, consortium level (<1000 nodes’) deployments.[4] For the purposes of this report it is important to note that a standard FL aggregation algorithm has been used called FedAvg which averages the parameters of each of the locally trained models in order to generate a global model at each federated learning cycle.[5]

Two tools were selected to enhance the privacy, transparency and robustness of federated learning systems for use in the KnowRisk platform:

1. **Model score cards for federated model reporting**
   (adapted for federated learning)

Model score cards for model reporting is an established tool for documenting and communicating crucial information about machine learning models to relevant stakeholders - in an effort to increase transparency and accountability while reducing the risks from information asymmetry and misuse of AI.[6]

The model card created is a living document that describes a machine learning model developed for KnowRisk and has been adapted for a federated learning context.

2. **Record on negative impact (RONI)**
   (adapted for building consortium trust)

**Reject on negative impact** (RONI) was initially proposed as a defence mechanism against various forms of model corruption and data poisoning attacks targeting federated learning systems.[7] A new adaptation of RONI: **record on negative impact**, focuses on the context of a small consortium of organisations for which automation of penalties might be unfavourable for consortium cohesion. Therefore, the output of RONI takes the form of active federated model monitoring, recording the impact of model updates from each of the participating parties (insurers) on the global model and it leaves decisions regarding thresholds for negative impact and penalties to the hypothetical parties themselves.

[3] https://en.wikipedia.org/wiki/Federated_learning
[4] https://github.com/digicatapult/dc-federated
[5] https://arxiv.org/pdf/1602.05629.pdf
[6] https://arxiv.org/pdf/1810.03993.pdf
[7] https://www.usenix.org/system/files/sec20summer_fang_prepub.pdf

These tools were intended to assist the progression from actionability, as determined by the ethics roadmap, to operationalisation with use cases from KnowRisk in mind. However, due to the early-stage nature of the KnowRisk project, this application of two applied AI ethics tools is only intended to demonstrate how practical tools can be a beneficial aspect of a wider integration of ethics processes and capacity building. Therefore, it is worth stating from the outset that there are a number of other tools that could be applied fruitfully to the KnowRisk project as part of a later stage of development and the application of any tools does not, in and of itself, make a project ethical.

This report will describe how the chosen tools were selected, adapted, used and evaluated. The team hopes  that this example of how tools can form part of a broader engagement with ethics will be a useful case study for others to learn from.

## Tool identification and methodology

> One of the challenges to responsible adoption practices is that whilst the greatest impact (at potentially the lowest cost) can be made right at the beginning of project development, this is also the time with most uncertainty in terms of project definition and scope.

The phase of the KnowRisk project covered by this report was indeed an early one - to build a proof of concept - in which the contributions from each collaborator were planned to cohere only towards the end.

Consequently, for pragmatic reasons, the focus of the work on the development and the evaluation of tools to facilitate responsible technology adoption was on the elements of the KnowRisk solution that were most accessible to the authors - the technical contribution from Digital Catapult (albeit with due consideration of the wider context of the KnowRisk project).

Digital Catapult's technical contribution to the KnowRisk project has been to build a proof of concept to demonstrate how federated learning can be used to harness private data from multiple parties to build better prediction models while maintaining the confidentiality of said data.

The particular prediction model demonstrated in this proof of concept analyses text from insurers' risk reports and identifies the risks and mitigations within them. The federated aspect involves training the model separately on private risk reports from (putatively) multiple insurers, before aggregating the results to gain better predictions than could be achieved from an individual insurer's data alone. This specific prediction task is a component of the overall KnowRisk solution. It should be noted that the consortium identified several other potential applications for federated learning within KnowRisk.

The tool selection methodology was to use the ethics roadmap, as detailed in the **Ethics report,** to identify areas of ethical saliency in relation specifically to the federated learning proof of concept; to identify tools that might assist in adhering to, or monitoring of, responsible technology practices in those areas; and then to select tools for further investigation.

The following selection criteria were used:

- Do tools exist?
- How well do they address the particular issue identified in the ethics road map?
- How mature is the tool or how readily can it be used?
- Can value be added by doing this evaluation?
- What functionality is required?

Further, consideration of the evaluation design (such as if any potential adaptation of the tool might be required, and cost and ease of implementation) has been taken into account.  The selected tools were then implemented and evaluated for their ability to meet the requirement to mitigate risks or enhance benefits.

# Tool identification and justification

**To identify tools that might help to meet the requirements for transparency, privacy and robustness in the federated learning setting, it has been necessary to identify sub-tasks and objectives, before identifying possible solutions to these more specific objectives.**

For example, the ability to avoid model bias is one element of transparency, which could be achieved through a combination of careful design, communication of the methodology, agreed standards for data collection, sharing of information about the characteristics of the training data and the ongoing monitoring of model outputs.

In the federated learning setting, the responsibility for good practice is complicated by the presence of multiple participants (as will be the case in any artificial intelligence supply chain). The group therefore mapped the specific tasks and objectives to where in the federated learning supply chain they arise or where they need to be addressed (Figure 1, right).

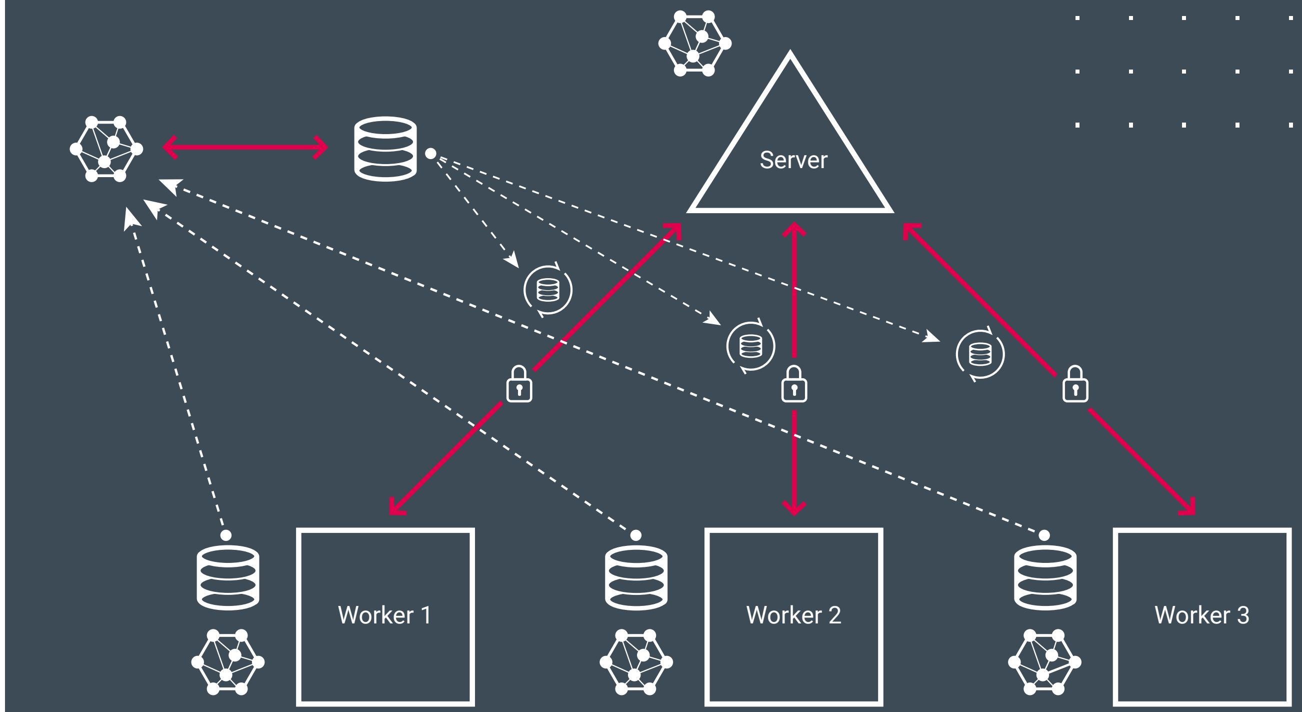


Figure 1: Sketch of federated learning system

This allowed research to focus on the potential tools that could assist in meeting these objectives.

The consortium used the AI Ethics Tools Typology,[8] a review of relevant literature and web searches, to identify candidate tools to evaluate against set criteria.

It quickly became apparent that tool selection in some areas was difficult or nonsensical during the very early stage of the KnowRisk risk prediction proof of concept, since choices depended on any future deployed implementation. Two examples of this are:

1. **The use of differential privacy to provide privacy guarantees.** There has been considerable investment in robust and secure implementations of differential privacy for machine learning (IBM[9] and OpenMined[10] for example), to a far higher standard (and trustworthiness) than could be achieved in the resource budget for this work. However, these implementations are typically dependent upon the machine learning model used, the machine learning framework deployed and the application (in choice of user-selected parameters epsilon and delta). Selection would be preferable later in the project timescale, when these elements are fixed.

   In the meantime, it is important to communicate that federated learning of itself does not offer any formal privacy guarantees. However, in most situations, the use of good regularisation techniques can avoid data leakage through memorisation and the effort required to reconstruct data from model updates currently makes privacy issues more a theoretical threat.

2. **Fairness and detection of bias.** Fairness has attracted a great deal of research interest and a variety of tools and approaches exist to help design fairer systems or identify biases post-hoc. Bias in the risk prediction model (as distinct from a KnowRisk system-level analysis of fairness) would arise from modelling (and scaling) existing biases in human processes or from imbalances or omissions in the training data. Given that the proof of concept was, by definition, a simplified case and one that used synthetically generated training data, it would not be meaningful to evaluate its fairness at this stage.

   Yet, there are distinct challenges to building and monitoring the performance of a federated system against fairness objectives that will need to be addressed at a later stage; not least is the challenge in understanding the characteristics of the training data used when it is distributed and strictly private. One promising approach to this latter problem is to derive synthetic data sets from the private ones, each with the same statistical properties.

[8] https://arxiv.org/abs/1905.06876
[9] Diffprivlib: The IBM Differential Privacy Library https://github.com/IBM/differential-privacy-library
[10] PyDP https://github.com/OpenMined/PyDP Openmined's python wrapper for Google's DP C++ DP library

## Transparency: selecting model score cards

As a priority, the consortium chose to focus on transparency since it is fundamental to the investigation of other areas of ethical saliency as well as to the specific collaborative aspects of solution co-development and customer onboarding. In addition, given the early stage of the project, focusing on transparency can identify any proof of concept limitations, such as the privacy and fairness examples discussed above, plus any requirements for further work.

Recent initiatives provide accurate information to participants and stakeholders on how the machine learning model has been designed, including its purpose and the data it relies on. These initiatives include: Partnership on AI's AboutML[11] project, an ongoing multi stakeholder initiative to enable responsible AI by increasing transparency and accountability with machine learning system documentation; IBM's AI Factsheets[12]; Google AI's Model Cards[13]; and Microsoft's use of Transparency Notes[14].

Therefore, in terms of the selection criteria has been based around the following:

a. **Do tools exist?** Yes
b. **Closeness of match with particular problem identified in the ethics road map(s):** These tools seek to increase model transparency through communicating facts and evaluations relating to their purpose, limitations and design. This should allow for informed decision making and facilitate trustworthiness across the supply chain and with customers and other stakeholders.
c. **Maturity of tool for use.** There is no established standard for documenting machine learning models, but there is much commonality amongst the proposed approaches. Some have been developed into toolkits or part-automated for use in certain circumstances. Integration into existing workflows is lacking, as is methodologies to continuously update the information as models are updated.
d. **Can the team add value by doing this PoC test?** We can pilot the use of model reporting in a federated setting and evaluate the ease of use and utility of the tools (against the transparency objective), and publish the results, adding to the know-how and templates

[11] Website: https://www.partnershiponai.org/about-ml/
[12] Original Paper: M. Mitchel et al, Model Cards for Model Reporting, 2019; Toolkit: Google AI Model Card Toolkit (2020).
[13] Original Paper: M. Mitchel et al, Model Cards for Model Reporting, 2019; Toolkit: Google AI Model Card Toolkit (2020).
[14] For example, https://azure.microsoft.com/en-gb/resources/transparency-note-azure-cognitive-services-face-api/

available. In the first instance, the primary audience for the specific model information will be other KnowRisk consortium members, to assist with the understanding and integration into the overall proof of concept. This prototype would inform the design of tooling to achieve transparency in the later, wider, context of deployment: i.e. insurers participating with their private data and other stakeholders in the KnowRisk system. The outcome will also be of interest to the wider machine learning and AI ethics community, as a case study for AboutML for example.

e. **Functionality required?** Identification and communication of required information relating to the federated risk prediction machine learning model.

Both model cards and AI factsheets are structured frameworks for reporting facts about machine learning models that have been proposed for widespread adoption. Both are in active development and refinement with various users, but neither has reached de facto standard use. Both have associated toolkits (although model cards has a tensorflow dependency) and they are similar in intent and content. The team made the decision to use model cards as the template for reporting as this initiative appears to have the most momentum.

> “Sometimes, the process of measuring or data collection can distort the thing it wants to measure; and sometimes data simply gets distorted at the entry point. The possibility of adversaries that feed data into the system with the explicit intent to distort the outcome can also not always be disregarded.”

## Robustness: selecting RONI

The second area of focus relates to the robustness of the risk prediction model, specifically to the observation in the roadmap that: *"Sometimes, the process of measuring or data collection can distort the thing it wants to measure; and sometimes data simply gets distorted at the entry point. The possibility of adversaries that feed data into the system with the explicit intent to distort the outcome can also not always be disregarded."*

In the case of distorted data, it is possible to make some adjustments to the data generation methodology for one or more workers to simulate distortions of data, to test the tools that can identify and mitigate against them. Such distortions include biases, data omissions, or processing errors. These are all harder to identify and mitigate in a federated learning setting and are worthy of further research.

There are a number of types of adversarial attacks that can occur within a federated learning system, for example:

- **Targeted model poisoning**: adversarial workers attempt to manipulate the training process in order to achieve specific aims - for example targeted misclassification while maintaining overall model performance, including stealthy poisoning to avoid detection[15].
- **Byzantine failures**: adversarial workers prevent the global model from converging on a reasonable optimum through the introduction of arbitrary model updates (random, drawn from a distribution with higher variance, and informed by knowledge of the system). Under normal FedAvg aggregation federated learning is not tolerant to even one adversary, so mitigations need to be introduced such as dynamically evaluating worker subsets during cost minimisation using Krum[16]. This approach converges in polynomial time and does not need a supplementary test/validation set beyond how the global model is already evaluated. An implementation of the Krum aggregation function is also implemented in IBM's Federated Learning Library[17].
- **Data poisoning, e.g. dirty label data poisoning attacks:** adversarial workers train local models on deliberately corrupted data in a targeted (where specific labels are changed) or untargeted manner (where labels are to some extent randomly allocated to decrease local model performance and therefore impact the global model)[18].

[15] http://proceedings.mlr.press/v97/bhagoji19a/bhagoji19a.pdf

[16] https://proceedings.neurips.cc/paper/2017/file/f4b9ec30ad9f68f89b29639786cb62ef-Paper.pdf

[17] https://github.com/IBM/federated-learning-lib/tree/main/examples

[18] https://arxiv.org/pdf/1712.05526.pdf

Overlap with general robustness measures:

- **Class imbalance and bias:** Given that local data is not directly observable, this makes efforts to counter class imbalance and potential bias difficult in an FL setting[19]. Tools that attempt to detect adversarial attack by looking at the effect of model updates on the global model may confuse updates using bad data with deliberate attacks.

Examples of defences that could be deployed to detect and mitigate some of these attacks include use of more robust aggregation algorithms: such as Krum or a trimmed mean method and local update monitoring systems, such as reject on negative impact (RONI); error rate based rejection (err); loss function based rejection (LFR); and a combination of the two used to reject local models[20].

Justification:

a. **Do tools exist?** Yes. There are aggregation function approaches to mitigating against Byzantine attacks and there are further approaches to local model poisoning attacks as well as techniques that can be adapted for federated model monitoring like RONI.
b. **Closeness of match with particular problem identified in the ethics road map(s):** In the context of cross-silo federated learning, a tool to monitor for a potential attack or failure can increase trustworthiness amongst users and the capacity to act. Given that federated learning systems are distributed and opaque with regards to data, a tool to measure and potentially take action based on per-worker/per-party model performance is a good match for the needs identified in the ethics roadmap around consortium governance and the auditing of privacy preserving machine learning systems.
c. **Maturity of tool for use.** IBM has quite a complete FL library that includes some implementations of adversarial robustness measures and there are many academic papers available. The use of Digital Catapult's federated learning library can make the adaptation of tools like RONI easier to implement.
d. **Can we add value by doing this PoC test?** It would be valuable to add robustness features to Digital Catapult's Federated Learning library. Furthermore, there is additional value in demonstrating the mechanisms through which organisations can collaborate with these tools, even if the tools themselves are not novel.
e. **Functionality required?** 1) Monitor the impact of updates from each party (worker node) on the global model 2) Provide actionable insight on potential attacks or failures to users through data visualisation or alert functionality.

There are a number of techniques that provide some level of defence or mitigation against adversarial attack such as the use of more robust aggregation algorithms, however in the context of the KnowRisk project, which is not dealing with defending a production federated learning system, the team decided that a monitoring approach would be a better fit for the needs of the project, as explored in the ethics roadmap. This led us to choose reject on negative impact (RONI) as a tool to adapt and implement with a focus on recording anomalies in performance on a per-worker/per-party basis rather than rejecting model updates automatically.

[19] https://arxiv.org/pdf/2008.06217.pdf
[20] https://www.usenix.org/system/files/sec20summer_fang_prepub.pdf

# Tool adaption and experimental design

## Transparency adapting for federated learning

There are a number of additional considerations that had to be factored in when adapting the Model Score Cards for Model Reporting tool for:

- **The overall KnowRisk project**

KnowRisk is an ambitious and mid-TRL (technology readiness level) applied research project in which the partners are working on interrelated but highly specialised components. The application of federated learning to a component of KnowRisk was crucial to a subset of partners (Intelligent AI and Digital Catapult), who were working on a Natural Language Processing (NLP) application for extracting insight from insurance and risk report documents. For the other partners it would serve as a useful template for applying similar approaches to other components of the project at a later stage. Together with the fact that the intended users (insurers) of the NLP model were not regularly engaged with the development process, the primary audience for the model score card was the KnowRisk consortium itself - particularly the partners directly involved in developing the machine learning model.

- **The specific machine learning use case**

As limited data was available for the proof of concept machine learning model, this necessitated the use of some synthetic data and restricted the scope of the work to producing a demonstrative proof of concept that could be built on at a later stage.

- **Federated learning**

The application of a federated learning approach meant that several new pieces of information needed to be provided on the model scorecard. For example: the number of worker nodes or parties; federated learning framework being used; and high level information on the data distribution between the parties if that was available. Furthermore, it was decided to document the use of the RONI tool in the model scorecard so the experimental setup for RONI is included.

The experimental implementation of the adapted model score card was to draft a version alongside the development of the federated learning model and share it with the consortium as part of the final project outputs to inform future work.

## Robustness: adapting for federated model monitoring

This work has been inspired by a number of sources but primarily a description of the reject on negative impact (RONI) tool for adversarial robustness detailed in *Local Model Poisoning Attacks to Byzantine-Robust federated learning*[21]. In order to implement a version of RONI which *recorded* rather *rejected* model update performance the team made changes to implementation of the FedAvg aggregation function, used in the Digital Catapult Federated open source federated learning library[22]. The changes enabled the recording of model performance with and without each worker update. The code for the RONI implementation will be published on the open source repository as part of the KnowRisk dissemination plan.

In the implementation of the NLP risk and mitigation classification model for KnowRisk this meant that for each federated learning cycle (when each worker update has been collected and aggregated into a global model) the RONI feature evaluated the performance of a model which only aggregated three of the four worker updates (a subset model). If one or more of the workers were consistently lowering the quality of the global model, compared to the other workers, then this could be seen in the relative performance metrics.

## Datasets:

The experimental setup for RONI involved creating three datasets which combined ground truth labelled data (risk and mitigation sentences), synthetic data, and “trash” data - which is irrelevant data designed to predictably lower the performance of the model that is trained on it.

1. **No trash dataset:** consisted of a held back test dataset of real risk and mitigation sentences. This dataset was used as a held back test set to evaluate the performance of the subset and global models.
2. **Equally corrupted worker dataset:** This dataset consisted of two thirds real risk and mitigation sentences and one third trash data. This data was split between four worker nodes.
3. **Unequally corrupted worker dataset:** This dataset is identical to the equally corrupted dataset except that one of the worker datasets is substituted for a dataset that is completely trash.

[21] https://www.usenix.org/system/files/sec20summer_fang_prepub.pdf
[22] https://github.com/digicatapult/dc-federated

It should be noted that in a production federated learning setting it may not be possible to have a globally shared test dataset (number 1 above) due to an inability to access the datasets directly. In practice however, it was anecdotally observed that obtaining limited access to illustrative datasets is common in federated learning consortia and is still easier than obtaining full access. Furthermore, using generative networks, it may be possible to generate a synthetic representative test set in a privacy preserving manner; an area certainly worth exploring in future research.

**For full dataset details** consult the model score card found in the Appendix.

**For modelling details:** consult the model score card found in the Appendix.

## Experiment:

A federated learning system was deployed with four worker nodes and a central server. The machine learning model was trained on local subsets of data before being aggregated by averaging the model parameters (FedAvg).

After each federated learning cycle, the RONI feature implemented would evaluate the performance, (in terms of classification accuracy) with respect to a held out (no trash) dataset, of each subset model (combinations of three out of the four worker updates) and the global model (an aggregation of all four updates).

To see if the RONI feature provided useful insights that could improve the robustness of the federated learning system, the team ran over 10  federated learning cycles for both the equally corrupted worker data set (2 above) and the unequally corrupted dataset (3 above).

When the performance curve (accuracy over FL cycles) is plotted for both datasets this should detect whether one of the subset models is consistently better than the rest (because it excludes the corrupted worker).

# Results

## Model score card for federated model reporting

See the appendix for the full model card.

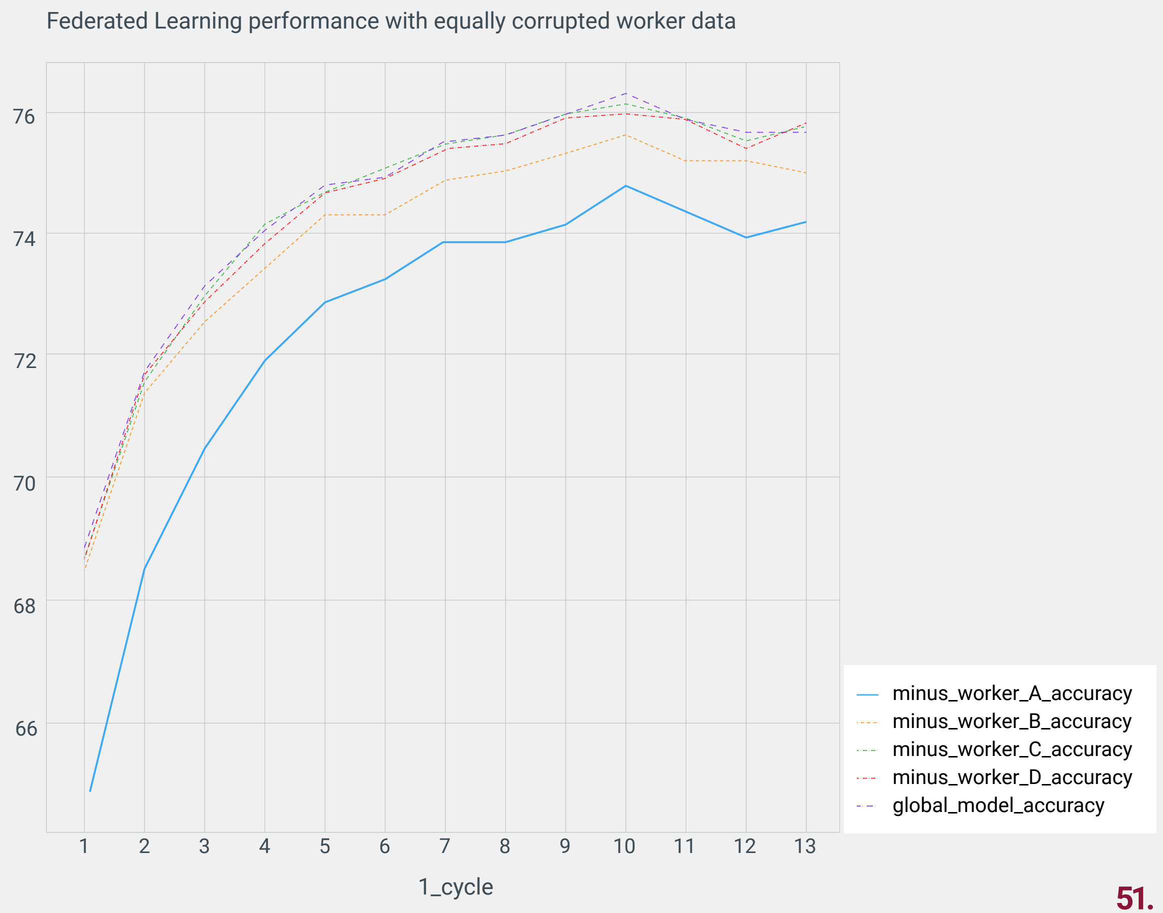
## Record on negative impact local (RONI)

In Figure 2 (right) and Figure 3 (below) the purple line represents the performance of the global model against the held-out test set and the other lines represent the performance of each subset model which exclude a specific worker in the aggregation.

In Figure 2, which shows the results for the equally corrupted dataset, you can see that no one subset model is consistently outperforming the other models with red and green showing similar performance to the global model (in purple). This doesn't indicate that excluding a specific worker increases performance.

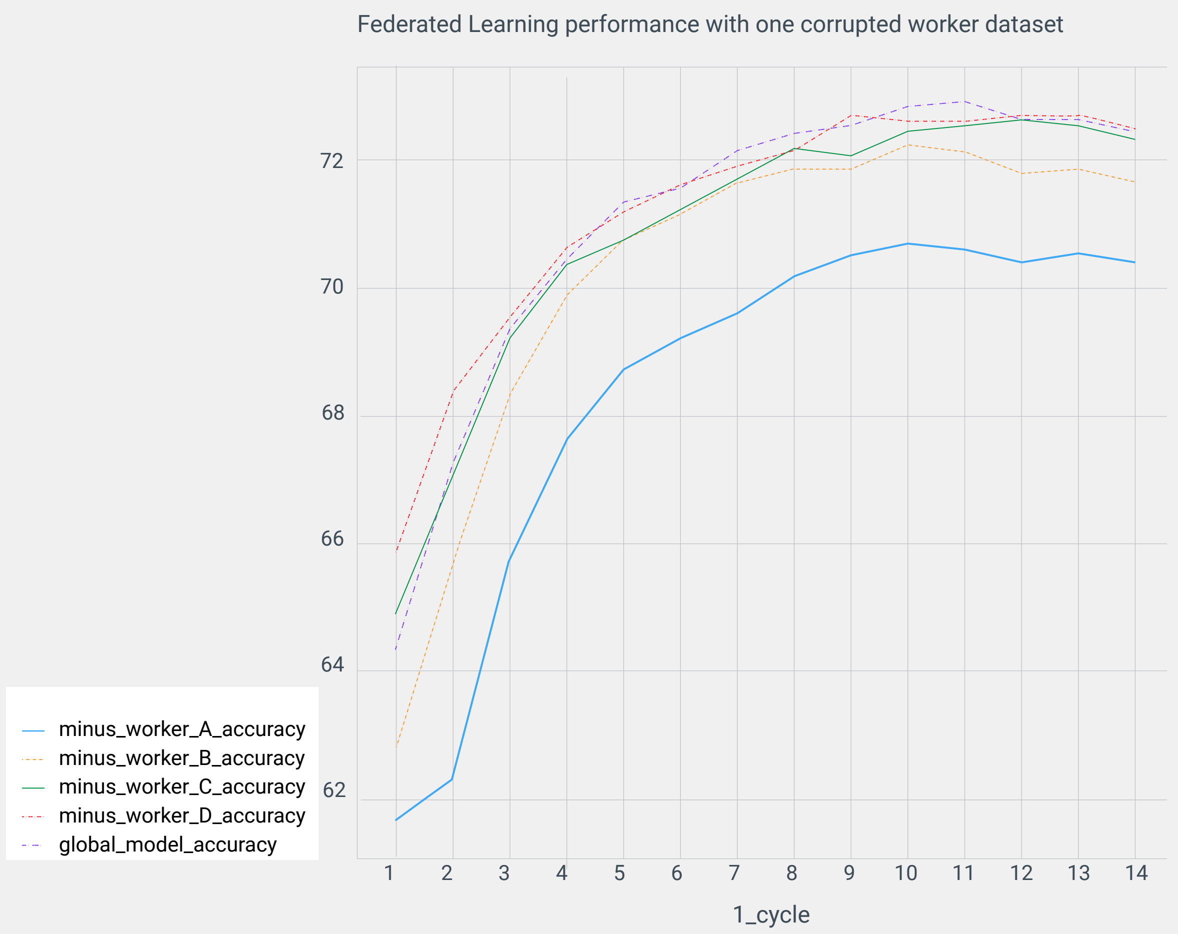
Below in Figure 3 you can see that the model subset represented by the red line (excluding worker D) is consistently outperforming the other model subsets and more closely tracks the global model. This indicates that worker D might be worth additional investigation for poor quality data or adversarial threats.

Figure 2: Performance (classification accuracy) on the y axis and federated learning cycle on the x axis for the **equally corrupted dataset**.



Federated Learning performance with equally corrupted worker data

Figure 3: Performance (classification accuracy) on the y axis and federated learning cycle on the x axis for the **unequally corrupted dataset**.



# Evaluation and discussion

## Transparency: model score card for federated model reporting

The main thing to note is that this model card does not conform to the template provided in Google's original paper or to any other template (such as IBM's factsheets). The model card needed to be tailored to fit the KnowRisk context - that of a proof of concept model unintended for production, with a relatively limited audience - as it was difficult to create a readable and meaningful document by adhering strictly to any template, which does not come as a surprise when even Google does not adhere to its own template in its published model cards[23].

**There is clearly a need to tailor model cards to different contexts, but that must be balanced by the need to create some consensus about what information a model card must contain in order to achieve its aims.**

The information in the model card results from an exercise of judgement rather than best practice (since these norms do not yet exist), and is subject to iteration and feedback with its intended audience to best achieve saliency and transparency.

The model has been produced as a snapshot pertaining to the final delivered model for the KnowRisk project. If it has value in communicating

[23] E.g. https://modelcards.withgoogle.com/object-detection

between the project participants at this point, that value will disappear as soon as the project evolves and new data, models and integrations occur. Furthermore, model cards (and similar) appear to be most applicable to situations with a single model owner, rather than a complex integration of models, potentially each originating from different owners' (as is envisaged in the KnowRisk consortium, or other collaborative projects); or where federated learning means that the information needed to complete a model card may itself be distributed or may need additional coordination and/or amalgamation.

It is clear that the use of model cards does not fulfil the requirements of transparency and responsibility in a federated setting, even without exposure to potential participants and users of the model (who are currently artificial). This is because the cards are currently static, it is not clear who is responsible for them, nor how to amalgamate distributed contributions. As the risk model is anticipated to be only one component of the KnowRisk solution, transparency is vital. In combination, the potential limitations and risks of the solution might become both more opaque and more acute. Therefore, a holistic communication of solution capabilities and risks will require integration of model cards (or similar) for each component.

Where possible, automating reporting tasks will help to integrate them into workflows. Since this is a common problem in machine learning, third party solutions and tools might assist - if they are sufficiently mature and flexible. One promising startup initiative is Parity (getparity.ai), a platform currently in beta for automating model card workflow, including the allocation of tasks and responsibilities, while maintaining flexibility so that users can specify the required information fields.

The resulting model cards must not only be accessible, but accessed too. The current model card will be published to a shared private github repository, which allows the information to be located alongside the model and associated with a specific version, at the very least, but it will not be the ideal choice for everyone. At some point, it may be suitable to make the model card public via the KnowRisk website and consider ways to encourage interaction with it, similar to Google's model cards.

## Robustness: RONI

Our experience of implementing and evaluating the record on negative impact (RONI) tool clearly shows that, even in a relatively controlled and artificial environment, the results are not concrete enough to inform meaningful alerts, let alone automatic mitigations (such as rejecting worker updates).

In a cross-silo consortium setting for federated learning, tools like RONI give the parties some means of monitoring model performance that may be indicative of corrupted data without actually needing to see the data itself. This insight is useful, but such tools can only be part of a wider set of socio-technical solutions - some of which are covered in the wider **Ethics Report**.

RONI could also form part of a suite of federated learning tools that include federated analytics tools that can answer questions about the statistical attributes of distributed datasets without breaching privacy. For such tools to be useful, it is recommended that they are introduced to participating parties early so that everyone is aware that the capability to detect potential attacks or failures is present. If this discussion is carried out as part of a wider conversation about ethics and accountability, as is occurring as part of the KnowRisk Ethics workstream, then tools like RONI can contribute to building overall consortium trust in the system. What is crucial is that the application of such tools is not done in a silo and is openly and clearly discussed with all the relevant parties.

# Conclusion

**This report is intended to serve as an open example demonstrating the consortium's planning and thought process in applying AI ethics tools, from the outset, within an early-stage collaborative research project.** It i s pleasing to see that the practical AI Ethics tools ecosystem has continued to grow; indeed, many resources have been added since the Digital Catapult tools survey in 2019[24]. As mentioned in the introduction to this report, the application of only two tools was never meant to be exhaustive but can serve as an illustrative example for the KnowRisk consortium and other practitioners.

**Regarding the specific tools selected, both model score cards for federated model reporting and record on negative impact (RONI) demonstrate some value as tools for increasing transparency and robustness.** However, it is also clear that in a production deployment of federated learning there is need for a whole system perspective to implement infrastructure to support effective model monitoring, privacy, robustness, accountability and appropriate consortium incentives. This socio-technical work should form part of a wider engagement with ethics, which is why this document should be read in the context of the **Ethics report**.

[24] https://link.springer.com/article/10.1007/s11948-019-00165-5

# Appendix

The consortium believes that this early and open-ended work in selecting, adapting, implementing and evaluating applied AI ethics tools at such an early stage of a project has been a valuable part of the overall Ethics workstream of KnowRisk.

Much of the essential ethics work of distilling and communicating ethical values, engaging in critical discussions, as well as interrogating potential risks and benefits has been covered in the Ethics report. By drawing a line from the outcomes of that work - from the agreed areas of ethical concern in the ethics roadmap, to specific tools that can operationalise ethics in the form of concrete procedures and processes, for example - this ethics tools work can help bridge the gap between the discussion of ethical issues and integrating ethics into the technology itself, as well as the human systems that operate it.

## Model Score Card for Federated Model reporting v0.2 - RONI experiments

Model Card Date: 01/06/2021
Model Version: 2

## Risk prediction

The risk prediction model has been developed by Digital Catapult for the KnowRisk project to classify risk/hazard and mitigation sentences in insurance risk reports.

- **Input:** Natural Language from insurance risk reports. Sentences that were either “risk” or “mitigations” were embedded in dynamically generated documents composed of the sentence of interest and “confounding sentences” from a variety of sources.
- **Features:** Words are first one hot encoded in an n-dimensional vector (where n is the vocabulary size) and then a learned embedding of size m. A document is hence a bag of embeddings. These are averaged to form a single input vector to input to the model for classification.

- **Output:** The model can classify whether a document contains a risk or mitigation. For each document input it returns a single label.
- **Model architecture:** Bespoke implementation of multi-class logistic regression over bag of word embeddings with the embedded dimension hyperparameter set to 10 using the PyTorch machine learning framework. The size of the model depends on the number of unique words in the corpus.
- **Training type:** federated learning.[25]
  - » The model is trained locally for 40 epochs at four (virtual) workers and the model updates are aggregated by a central server before being shared back to the nodes for further local training.
  - » A simple (FedAvg) average aggregation model is used in which model parameters are averaged across all nodes
- **Monitoring:** A record on negative impact (RONI[26]) feature has been implemented to evaluate the impact of each worker update using a held back validation set. Administrators can set an alert threshold which detects significant irregularities and flags workers for further investigation. The updates are not automatically rejected as in the original reject on negative impact.

## Intended use

- The model is a proof of concept. It will form one component of a solution that automatically extracts risk and mitigation sentences from insurance risk reports and predicts a risk score for a specific building or set of buildings.
- The intended users are the KnowRisk consortium partners for the purposes of developing and demonstrating a proof of concept application.
- Use by any party other than a KnowRisk consortium partner, for commercial deployment, or use for risk assessment other than specific commercial property insurance related risks is not intended.

## Training data

Training data was generated dynamically by creating documents consisting of multiple sentences, each in the form of a tokenised list of words. One of the sentences was a risk/hazard or mitigation sentence while the remaining sentences were drawn from confounding sources such as a wikipedia dataset. Each document, as a whole, was labelled as a risk or mitigation depending on the label of the risk/mitigation sentence.

[25] McMahan et al, Communication efficient learning of deep networks from decentralized data, 2016
[26] https://www.usenix.org/system/files/sec20summer_fang_prepub.pdf

The sentences contained in the document datasets used in the RONI experiments were from four sources:

**Verified labelled data**: Ground truth correctly labelled “risk” and “mitigation” sentences shared by insurance partners (321 risk sentences, 389 mitigation sentences).

**Augmentation labelled data:** Augmenting sentences selected, based on similarity to ground truth data (cosine similarity on BERT embeddings), from publicly available residential risk reports via data.gov.uk[27] (679 risk sentences, 611 mitigation sentences). These sentences were given the label of the ground truth sentence that it was most similar to.

**Confounding data to generate input documents**: 5000 sentences from the Wiki-Split[28] dataset were used to pad out the document with confounding sentences.

**Simulated corruption data (“trash”) for RONI experiments**: 1000 sentences from the Large Movie Review Dataset v1.0[29]. These sentences were given random labels.

Combinations of these datasets were used to create more or less corrupted datasets distributed over 4 workers for the purposes of testing RONI:

1. No trash dataset: consisted of a held back test dataset of real risk and mitigation sentences. This dataset was used as a held back test set to evaluate the performance of the subset and global models. **(only contains verified labelled data or augmentation labelled data)**
2. Equally corrupted worker dataset. This dataset consisted of two thirds real risk and mitigation sentences and one third “trash” data.
3. Unequally corrupted worker data set. This dataset is identical to the equally corrupted dataset except that one of the worker datasets is substituted for a dataset that is completely “trash” **(only containing simulated corruption data).**

## Evaluation data

Evaluation data is generated in the same way as the training data. The test-train split is 90% test and 10% train (to offer a harder problem in this artificial experiment).

A distinction should be made between test data that is used in the training of local models, which is split from locally available data and the held back dataset that is used to evaluate subset model performance as part of RONI.

[27]https://data.gov.uk/search?q=fire+risk+assessment&filters%5Bpublisher%5D=&filters%5Btopic%5D=&filters%5Bformat%5D=&sort=best
[28] https://github.com/google-research-datasets/wiki-split
[29] https://ai.stanford.edu/~amaas/data/sentiment/

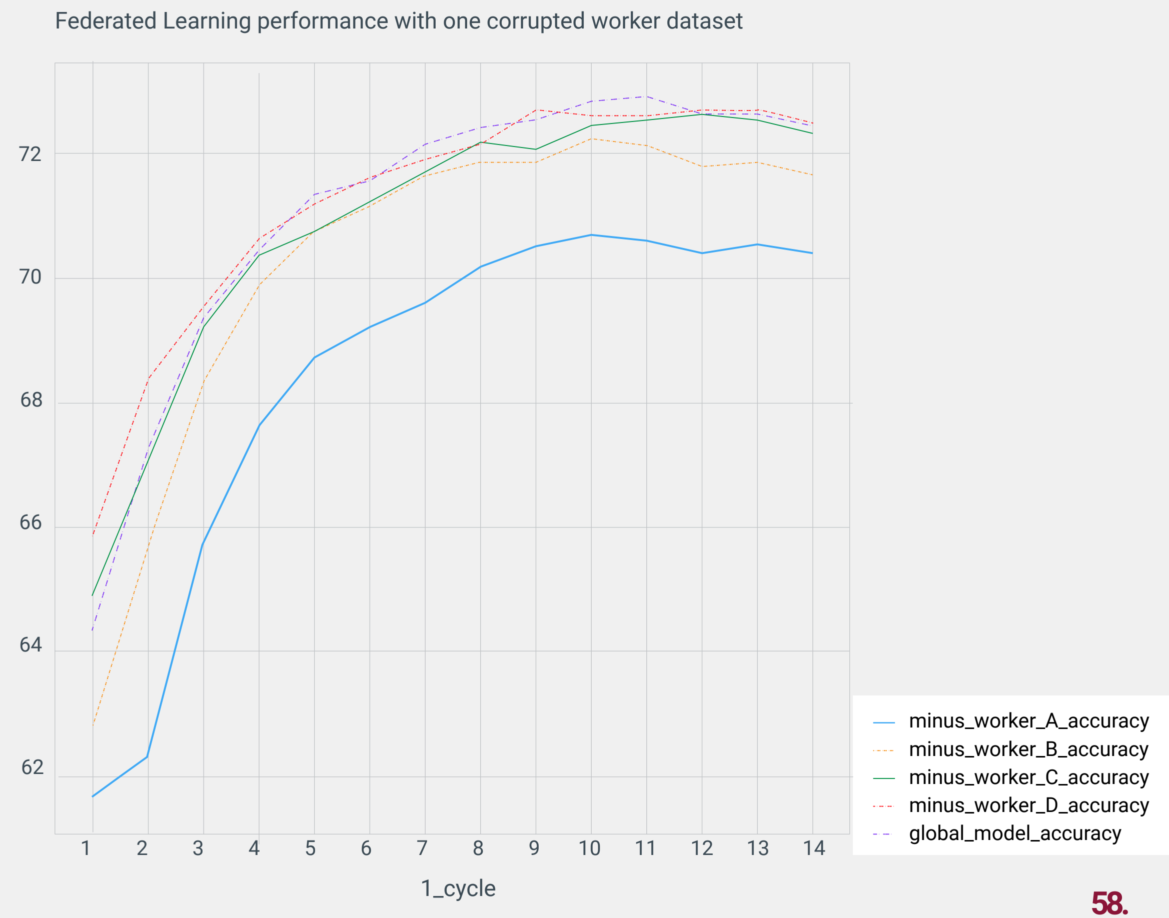
## Performance

Model performance is measured using classification accuracy against the evaluation data. Evaluation data is held both locally (normal for federated learning) and centrally (not typical for federated learning but necessary to test RONI).

Local evaluation is recorded every time a new global model is sent to each node and is tested after 40 epochs of local training.

Evaluation of subsets of the global model (excluding one worker to measure impact) are recorded at each federated learning cycle (once all four nodes have sent their local model updates).

Graph of classification accuracy over federated learning cycles:



## Limitations and ethical considerations

- This model is trained on partially synthetic data which is based on a limited number of risk reports. This model is not intended for use in production as it is not trained on suitably representative data. Real use cases might have much bigger vocabulary and / or classes which lead to bigger data requirements, models and consequential training and deployment challenges.
- The choice of classes has been identified by KnowRisk consortium members to be relevant for a proof of concept, but these may not be a complete or entirely appropriate set for real users and their data. In particular, the problems arising from class imbalance have not been investigated.
- Federated learning is intended for use in situations where individual contributors wish to keep their data private and secure, but it does not guarantee privacy. There is a risk that data leakage can occur through data memorisation or through reconstruction from model weight updates. In production, use of a range of techniques and technologies are recommended alongside conventional federated learning, such as: good regularisation strategies; differential privacy; secure multi-party computation; and homomorphic encryption.
- As with any machine learning model, the training data may contain biases, errors or imbalances that can impact on the efficacy and fairness of the model. In the federated learning setting, it might be more difficult to monitor and mitigate against these concerns as the underlying data from each worker is private. For production, further work is required to address these concerns.
- The performance of this model could have a significant impact on the aggregate risk scores that will be generated by the KnowRisk application. Therefore, there is potential for harm via biased selection or omission of risk/mitigations which then feed into a biased risk score, the purpose of which is to inform decisions regarding insurance claims. The monitoring and performance metrics for a deployed solution will differ from the simple classification accuracy used here and require further thought.
- This model is intended as both a proof of concept machine learning model to offer specific functionality to the KnowRisk platform (classification of risk and mitigation sentences) and a demonstration of the selected AI ethics tools in action including this one (model score cards for federated model reporting) and record on negative impact (RONI) and so should be viewed as demonstrative and not ready for production use.

# KnowRisk: CONSTRUCTION

# KnowRisk: CONSTRUCTION

## Insights from policy engagement sessions

**Although several UK industries have suffered from stagnating productivity levels in the past decade, construction is perhaps one of the hardest hit. This industry is relatively volatile and particularly sensitive to levels of confidence of both consumers and businesses, as well as fluctuations in economic rates.**

As the proposed KnowRisk solution utilises advanced technologies to protect against supply chain risks and disruption, policy engagement sessions were held, with attendance from key stakeholders in the UK construction industry, as well as government departments, industry bodies and technical advisers, to determine the ways in which using advanced digital technologies could benefit industry at large and any potential barriers to adopting these technologies.  Although a small focus group was consulted, insights were presented that reflect the needs across industry, spanning across company size, geographical location, position in the supply chain and other relevant factors.

The policy engagement sessions conducted on this topic allowed peers to share thoughts and ideas, building on common ground and new insights, resulting in several findings that will help the KnowRisk project to provide a solution that is most beneficial to industry.

The key findings from the policy engagement sessions were as follows:

### Top Risks - these can be categorised into the following groups:

- **Contracts and procurement**
- **Payment practices**
- **Late completion of projects**

### Barriers to adoption

- **Culture, knowledge and skills**
- **Industry relationships**
- **Return on investment**
- **Current technology and innovation landscape**
- **Business capacity**

### Risks associated with construction supply chains

- **Contracts and procurement**
- **Payment practices**
- **Late project completion**

# A background to supply chains in construction

By nature, the construction industry is heavily reliant on supply chains and any change in the chain can have a significant effect on the completion and profitability of products. As noted previously in the Weather Ledger project, contractors in construction supply chains are increasingly faced with both higher material costs and falling order numbers, which in turn squeeze contractors' margins and an issue in one area could have a significant impact on trading.[1]

This has become particularly apparent in recent years, with increased inflation and product scarcity, caused by a combination of factors related to COVID-19, Brexit and the Suez Canal blockage in March 2021, all of which contributed to significant disruptions in UK construction supply chains. The impact of these factors, the risks of which were not previously predicted to the scales that they reached, meant that the disruption of supply chain flow led to time losses, significant monetary losses, lost contracts and in some instances brought projects to a halt. Industry stakeholders have noted that the impact of some of these issues are expected to continue to have a lasting impact on supply chains for years to come.[2] The impact of Covid-19 on construction supply chains can be seen to have had the biggest negative effect on the construction industry, leading to an immediate 40% fall in growth in March 2020.[3]

However, construction supply chains are beginning to bounce back, with supply chain leaders working together to bring the industry back to its pre-pandemic levels. The construction industry's UK output in March 2021 was 2.4% (£334 million) above the February 2020 pre-pandemic level; repair and maintenance work was 7.7% (£377 million) above this level; while new work was 0.5% (£44 million) below. Other promising figures from consulting firm PwC[4] have shown that the construction sector has also made one of the strongest recoveries: its change in GDP between April 2020 and October 2020 was 70%, compared to just over 20% for services.

As the industry continues to pick up and look for ways to reduce risk to achieve greater levels of productivity and efficiency, supply chain leaders are considering a range of measures to accomplish these goals.

[1] Turner & Townsend, Q3 2019 UK Market Intelligence, https://www.turnerandtownsend.com/en/perspectives/uk-market-intelligence-q3-2019/

[2] Digital Catapult Policy Engagement sessions, May 2021

[3] https://www.ons.gov.uk/businessindustryandtrade/constructionindustry/bulletins/constructionoutputin-greatbritain/december2020

[4] https://www.ons.gov.uk/businessindustryandtrade/constructionindustry/bulletins/constructionoutputin-greatbritain/december2020

# Digitalisation in construction

One such measure is that of increasing the use of relevant digital technologies such as AI and machine learning in their supply chains,[5] as well as other industrial digital technologies that may help to improve profit margins and build stronger degrees of predictability, safer projects,[6] and more visibility[7] throughout the supply chain.

**Backed by government measures and policy, the construction industry is beginning to increase its openness towards advanced technologies and their use within supply chains.** Initiatives such as the 2018 Construction Sector Deal, which references the need to support a construction sector that is increasingly based on digital and manufacturing technologies, or the 2017 Made Smarter Review, which set out a vision for growth and increased productivity across the manufacturing sector through industrial digital technologies (IDTs), are part of a growing movement in the UK to use both advanced and elementary digital technologies to make construction supply chains more efficient and more productive. Additionally, programmes such as building information modelling (BIM) are becoming increasingly used in business and have been called to be used in all construction projects.

The UK government has previously mandated that by 2020, all construction projects should have incorporated BIM into their operations, in order to drive digital transformation within the industry through its Digital Built Britain programme.[8]

However, despite these initiatives and an attempt at digitalisation by industry, the construction industry remains widely regarded as one of the most traditional sectors of the UK economy, meaning that the take-up of digital technology in supply chains is particularly low and slow in progressing. Some industry experts have noted that part of the issue lies within the fact that there appears to be no concrete mechanism, policy instrument or other regulatory regimes by which the adoption of BIM and other relevant technologies can be enforced.

[5] https://www.ukconstructionmedia.co.uk/features/benefits-ai-construction/
[6] https://www.bimplus.co.uk/explainers/where-will-construction-gain-most-technology/
[7] https://www.ibm.com/downloads/cas/GJOMQOWL
[8] https://www.gov.uk/government/news/launch-of-digital-built-britain

# Risks associated with construction supply chains

While the risks associated with construction supply chains are numerous and continually evolving, as building practices, technology and other factors change, the top risks to supply chains cited by industry can largely be categorised into the areas of **contracts and procurement**; **payment practices**; and **late project completion**.

## Contracts and procurement

One of the primary risks cited as an issue to construction supply chains both domestically and internationally is contracts and procurement. Industry experts have noted that how contracts and procurement are currently set up means that in a large percentage of supply chains, there is limited visibility of various parties making up the chain, meaning that it is near impossible to effectively organise the supply chain from end-to-end.

Also noted by industry advisors, is that when setting terms of the contracts, some negotiators focus on the overall return of investment of that project and do not provide allowances to ensure that a construction project is realistic. It has been suggested that often more consideration is given to setting terms that ensure the provider will perform a service in the most rapid and cost- efficient way. While this consideration is important, it does not always translate into setting effective timelines and efficiencies.

Identifying the quality of a supplier in the procurement process has been raised as a significant risk in construction supply chains. When deciding between new suppliers - who may all offer attractive terms, such as ideal pricing and the promise of high-quality work - a frequent lack of quantitative data to support these offers mean that those procuring do not always have a method to choose the ideal supplier.

## Payment practices

Late payments not only disrupt the flow for the entire supply chain but can restrict construction growth.

One issue that has become increasingly discussed, particularly in recent years when some high-profile cases caught the attention of the media, is late payments in the supply chain. This is a problem that not only disrupts the flow for the entire supply chain but can adversely affect small and medium-sized enterprises (SMEs) much more than their larger counterparts.

An industry standard for regulating payment practices exists, in the form of the Prompt Payment Code - under the guardianship of the Department for Business, Energy and Industrial Strategy - which seeks to set a standard for payment practices across the industry to ensure that larger construction companies pay their suppliers on time. However, industry voices have previously noted that this code is voluntary, so smaller businesses do not always have sufficient legal recourse to help them receive payments.[9] Industry analysis on the subject found that, although the Prompt Payment Code was introduced in 2012, in 2019 contractors still paid their suppliers within an average of 43 days. It was also found that more than a quarter of invoices were not paid according to their original terms.[10] It should be noted that the Prompt Payment Code was updated in early 2021 to strengthen the standard and reduce the payment timeline for payments to small businesses. However, the code remains voluntary and as such, session attendees have raised concerns over the effectiveness of the recent update.

For some industry experts, this issue has been raised as particularly concerning, as it was suggested that late payments in construction supply chains restrict construction growth.[11] Late payments by construction corporations force many SMEs to pay their own suppliers late; they may be unable to fulfil planned investment intentions; and often need to acquire bank loans to manage cash flow. These delays can cause even more issues when parties owing payment cease trading, such as in the 2018 collapse of British multinational construction firm Carillion, which at the time of its folding was said to owe $7 billion in debt,[12] with suppliers who had provided services for them being left empty-handed.

## Late project completion

In 2001, it was estimated that 70% of government construction projects were delayed.[13] Two decades later, drastic improvement is hard to see, with delays reaching 65% in 2015[14] and the COVID-19 pandemic further exacerbating this, with an estimated 4,500 projects being delayed in May 2020. The late completion of projects often has a knock-on effect on other aspects of the construction supply chain, leading to forced delays on other projects, cost overruns and other issues that are not always covered by compensation.

[9] https://constructionmaguk.co.uk/late-payment-of-smes-in-construction-how-simple-legislation-can-solve-the-problem/
[10] https://www.constructionnews.co.uk/archive/cn-payment-100-the-best-and-worst-payers-revealed-25-03-2019/
[11] https://rlf.co.uk/late-payment-restricts-construction-growth/
[12] https://www.theguardian.com/business/2020/jan/15/carillion-collapse-two-years-on-government-has-learned-nothing
[13] https://www.designingbuildings.co.uk/wiki/Delays_on_construction_projects
[14] https://www.architectsjournal.co.uk/archive/almost-two-thirds-of-projects-were-late-in-past-12-months

# Barriers to adoption

Insights from session attendees revealed a multitude of further barriers surrounding digitalisation in construction supply chains. These reasons can broadly be categorised into the areas of **culture, knowledge and skills**; **industry relationships**; **return on investment**; the **current technology and innovation landscape**; and **business capacity**.

## Culture, knowledge and skills

One of the most frequently cited reasons for the lack of digitalisation within construction supply chains is a poor understanding of risk, coupled with a general lack of openness towards aspects of innovation.

With the industry being traditionally risk-averse, even more so in the recent uncertain economic landscape, traditional methods of construction, communication and operations have continued to dominate, with most industry stakeholders sticking to what has previously worked and some being reluctant to consider more advanced digital technologies.

For those in industry who are open to change and the adoption of innovative uses of technologies to improve their operations and external supply chains, a general lack of understanding of the technology areas, what the first steps should be and how to engage to best serve their needs are common barriers to adoption. Attendees of the policy sessions noted the lack of digital expertise within the industry as a significant issue, as a general understanding of everyday technologies does not necessarily translate into understanding the underlying requirements and potential capabilities of advanced industrial digital technologies.

Linked to this is a perceived difficulty throughout the industry to both attract and develop the right mix of skills and capabilities within the construction supply chains. With technological skills in high demand across the economy, both domestically and internationally, the competition for the most skilled and valued is particularly high. Having to compete with an entire industry in terms of brand recognition, salary and other metrics means that those in the construction supply chain, particularly SMEs in the process, may be unable to attract those that have the potential to enact transformative change throughout supply chains. Skills may also be an issue internally to construction companies too, as existing employees may not have the skills required to start the process of technology adoption and companies may not have the resources or capacity to upskill staff.

## Industry relationships

Industry relationships and the strength of the supply chain as a whole are key to assess the potential of digitalisation to rescue supply chain risk in construction. Industry experts have noted that whilst some parts of the supply chain have robust digitised systems and are confident in their abilities, the same level of trust does not exist across the supply chain.

Industry experts have noted that bringing the supply chain on the journey of digitalisation is at times the most difficult aspect, with issues of interoperability including differing standards; differing IT policies; and other similar issues, often acting as a barrier to cohesive and robust digital supply chains and upholding the perceived fragmentation of the supply chain. This issue is one that can be seen as so fundamental that one industry expert noted that *"It's great for the leaders to be leading but the market only functions as well as its weakest supplier."*[15]

## Return on investment

Discussions within industry have highlighted that the adoption of advanced digital technologies does not always translate across the entire supply chain because, as mentioned previously, players in the construction industry are faced with increasingly tight profit margins, meaning that investment intentions have to be very carefully considered, in order to be able to to keep afloat, let alone make profit. Industry experts have noted that few construction firms are willing to be disruptive or to invest additional resources in R&D, choosing instead to focus on aspects of continuity and survival. Because of this, those in the construction supply chain are forced to consider the return on investment of any new technological practices.

A broader uptake of general project bank accounts (PBAs), which see members of the construction supply chain receiving payment in five days or less from the due date, easing cash flow through the system,[16] has been suggested by session attendees. This process ensures that different parties throughout the supply chain are sufficiently reimbursed for their investment and enable the delivery outcomes that arise as a result of the technology adoption.

## Current innovation and technology landscape

For those in the construction supply chain, issues surrounding the availability and features of the current technology landscape, and accessibility of the innovation landscape at large, often arise in conversations surrounding technological adoption. One such concern is that the current provision of tools aimed at solving a particular problem, gap or improving

[15] Digital Catapult Policy Engagement sessions, March 2021

[16] https://www.gov.uk/government/publications/project-bank-accounts

productivity is not up to the required standard to gain sufficient value. While industry experts note that this is not the case across the board, there is an understanding that several products marketed to parties across the supply chain do not sufficiently offer the value that the vendors have promised them as having. One example given is that some technological solutions offering valuable, reliable and up-to-date data analysis do not provide the data in an accessible format as required by suppliers. More specifically, some experts call for better tools for industry, including the development of automatic verification and validation tools.

Session attendees have noted that these barriers can be linked to price. Often, the products that have the most value to industry are at a price point that is inaccessible to many parties in the supply chain, often disproportionately affecting SMEs in the chain. Some in industry note that solutions appear to be designed for larger organisations, not keeping SMEs in mind in terms of affordability, ease of use and requirement of skills.

The interoperability of new technologies with legacy IT systems in construction, whether in the form of add-ons or new equipment, is also a commonly cited issue within industry, with some asserting that it is critical to construction productivity with some even calling this issue the industry's 'silent killer'.

## Business capacity

Construction SMEs are more likely than others within the supply chain to struggle between balancing the scouting and integration of new digital technologies and managing day-to-day operations as they focus on keeping afloat and deal with the often more significant impact of low profit margins.

This barrier could particularly be an issue as the longer this lack of capacity due to time restraints continues, the longer SMEs are likely to fall further behind their larger counterparts in adopting IDTs, which could potentially erode both domestic and international competitiveness.

# Conclusion and recommendations

The diverse risks, challenges and many of the barriers to adoption of technology were determined to come from a root cause in the construction supply chain: **a disconnected supply chain.**

The opportunity exists to address this disconnect and identify a path through which parties throughout the supply chain, regardless of tier, company size, sub sector or geographical location, can better communicate to reduce risk and strengthen their levels of productivity and profitability and competitiveness in the global economy.

Recommendations from industry experts include that, as the main client of construction, the UK government should mandate that BIM and other construction-related advanced digital technologies are enforced and monitored.

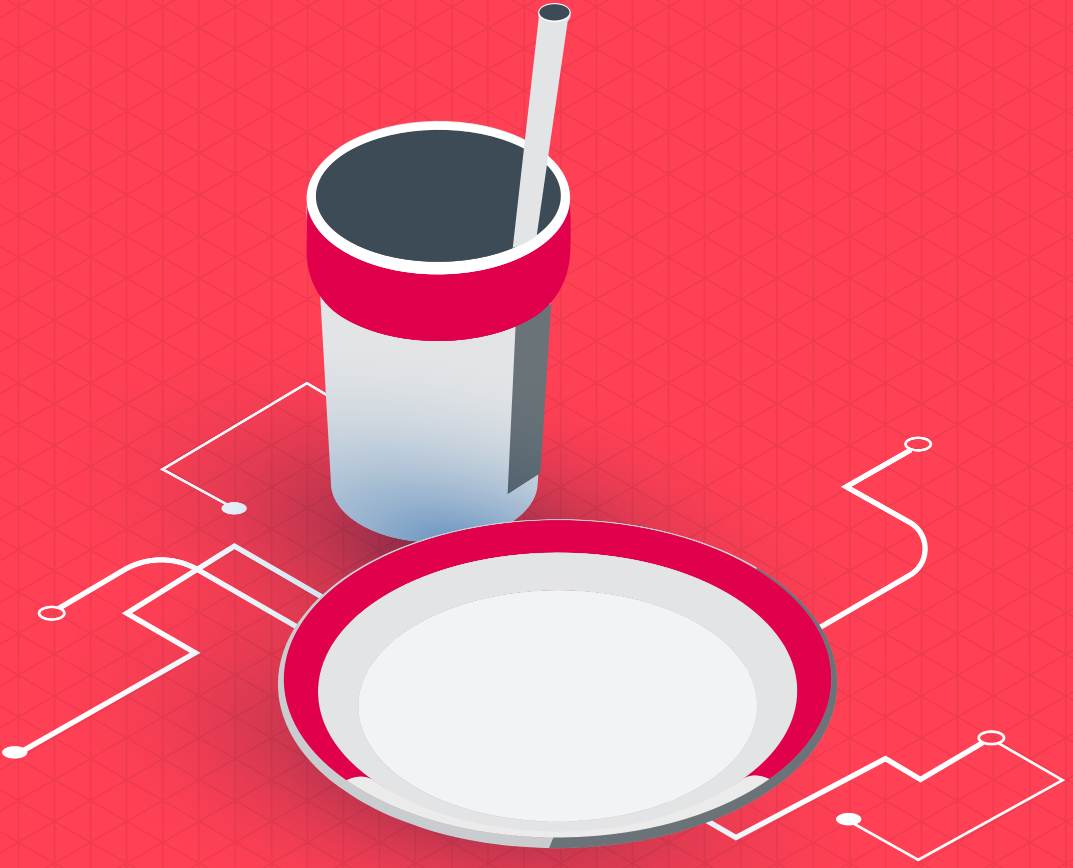
With regards to the proposed KnowRisk solution, industry experts consulted in the engagement sessions noted that the solution could be useful for the following reasons:

- **Visibility**: the solution could allow for extended visibility of the supply chain.
- **Collaboration**: the KnowRisk advanced technology solution would increase collaboration within industry, addressing the disconnect between various parts of the supply chain.
- **Expectation management**: the KnowRisk solution could offer various parties an improved overview of supply chain operations and therefore manage expectations more effectively.
- **Sustainability**: the solution opens up new avenues for increasing sustainability across supply chains.

Other interventions will be required to fully combat the identified risks and challenges, ranging from policy-based and regulatory solutions to other industry-led and potentially academic-led interventions.

Conversations with industry experts have revealed that for there to be significant change, a joint push from all of these players in the construction ecosystem is required to successfully leverage the high technological capacities and rich ecosystem that the UK has to offer. Experts believe that with the industry's driving force and broad expertise, the UK can become a leading economy for a construction industry that embraces innovation and efficiency.

# KnowRisk: FOOD & DRINK



**KnowRisk:**

# FOOD & DRINK

## Insights from policy engagement sessions

In the last decade, what constitutes a supply chain risk has changed considerably. In 2010, supply chain leaders considered the most significant risks and sources of exposure to be mainly economic, concerning raw material price fluctuation, currency fluctuations and market changes, along with energy and fuel price volatility.[1] Prior to the COVID-19 crisis, food and drink supply chain leaders were already raising issues related to food scarcity in the face of exponential population growth.

As the proposed KnowRisk solution utilises advanced technologies to protect against supply chain risks and disruption, policy engagement sessions were held, with attendance from key stakeholders in the UK food and drink industry, as well as government departments, industry bodies and technical advisers, to determine the ways in which using advanced digital technologies could benefit industry at large and any potential barriers to adopting these technologies. Whilst a small focus group was consulted, insights were given that reflect the needs across industry, spanning across company size, geographical location, position in the supply chain and other relevant factors.

The policy engagement sessions conducted on this topic allowed peers to share thoughts and ideas, building on common ground and new insights, resulting in several findings that will help the KnowRisk project to provide a solution that is most beneficial to industry.

This session fits into a wider conversation about challenges facing the food and drink industry and will be followed by further workshops and activities separate to the KnowRisk project, scheduled to address these challenges..

1https://www.pwc.com/gx/en/operations-consulting-services/pdf/pwc-and-the-mit-forum-for-supply-chain-innovation_making-the-right-risk-decisions-to-strengthen-operations-performance_st-13-0060.pdf

The key findings from the policy engagement sessions were as follows:

**Top Risks** - these can be categorised into the following groups:

- **Changing trends:** e.g demand-driven change in consumer behaviour
- **Natural and geopolitical:** catastrophic events, such as climate change and trade wars
- **Regulatory:** litigious stakeholders or changes in regulatory environment

**Potential benefits of KnowRisk** - attendees noted that the KnowRisk solution could allow for:

- **Extended visibility of the supply chain**
- **Aid collaboration**
- **Expectation management**
- **Opportunities for sustainability**

**Policy and industry solutions** - could include:

- **Supporting government's understanding of food and drink challenges**
- **National systemic risk assessment**
- **Agreed guiding principles**
- **A collective voice**

# The changing face of risk in food and drink supply chains

As recent research by the World Economic Forum illustrates,[2] the top perceived global risks by impact are societal - with infectious diseases and livelihood crises as most important, closely followed by environmental risks - including climate action failure, biodiversity loss, natural resource crises and human environmental damage. The risk of economic shocks are perceived to be decreasing both in frequency and impact on supply chain cost and performance.

Between 2020 and 2021, the risk profile, as perceived by supply chain leaders, changed significantly. UK food and drink supply chains, like many industries in the UK, have encountered disruptions, such as the COVID-19 pandemic, the March 2021 Suez Canal blockage and the UK leaving the European Union. The combination of these unprecedented shocks has increased the need and urgency to assess supply chains.

In 2020, 92% of UK CEOs[3] said that the disruptive impact of the pandemic forced them to rethink their global supply chain - a figure that represented a higher margin than any other surveyed country. In addition, the Suez Canal blockage, one of the most significant examples of logistical disruption in recent history, demonstrates how one single point of failure can have a ripple effect on a number of industries. This event resulted in significant economic disruption as well as many downstream production line delays and blockages.

[2] http://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2021.pdf

[3] https://home.kpmg/uk/en/home/insights/2020/09/uk-ceo-outlook-pulse-survey-2021/the-no-1-risk-to-your-business-your-supply-chain.html

# Mitigating risk and increasing resilience

The need to re-evaluate supply chains translates into mitigating risk and increasing resilience. More generally, a 2020 McKinsey survey[4] of manufacturing and supply chain professionals found that 93% plan to focus on making their supply chain more resilient given the challenges brought about by the pandemic: for instance, in the food and consumer-goods industries, 100% of respondents experienced production and distribution problems and 91% had problems with suppliers. This has been due to companies having very little time to address logistics disruptions, product shortages and abrupt shifts in demand. Similarly, according to KPMG's 2021 CEO Outlook Pulse Survey[5] from early 2021, executives from leading global companies perceived supply chain risk to be the third greatest risk to company growth over the next three years, rising from the eighth spot in February 2020.

Many of these concerns arise from the fact that manufacturing industries have been accustomed to efficient, rather than resilient, operations. Resilience is typically built on two conceptual notions: functional redundancy (no single point of failure within the chain) and diversity (in geographies, products and food suppliers).[6] Typically, both notions are incompatible with efficiency, which relies on the just-in-time supply chain concept (particularly, with food and drink), with preferred suppliers and limited scope for product substitution.

In May 2020, Gartner conducted its Weathering the Supply Chain Storm survey[7] which confirmed that only about 20% of supply chain leaders believed their supply chain to be highly resilient in terms of its ability to respond effectively to changes in trading conditions. This report explores exogenous factors, coupled with a series of new challenges for food and drink supply chains. These challenges including: changing consumer behaviour, such as the increased demand for fair trade and sustainable products; the growing trend of online and last-mile deliveries; and workforce shortages,[8] create a number of complications and risks for food and drink supply chains. The use of industrial digital technologies and solutions can help address and alleviate some of these risks.

[4] https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/Operations/Our%20Insights/Resetting%20supply%20chains%20for%20the%20next%20normal/Resetting-supply-chains-for-the-next-normal.pdf
[5] https://home.kpmg/uk/en/home/insights/2020/09/uk-ceo-outlook.html
[6] https://committees.parliament.uk/writtenevidence/7496/html/
[7] https://www.gartner.com/en/supply-chain/trends/weathering-the-storm-supply-chain-resilience-in-an-age-of-disruption
[8] https://www.fdf.org.uk/fdf/resources/publications/reports/covid-19-impact-on-food-drink-manufacturing/

# Digitalisation in food and drink supply chains

**Digitalisation presents many opportunities for the food and drink sector, including improved planning and forecasting; lower costs; and a boost in productivity, efficiency and profits.**

Traditionally, and in comparison to other UK industries, the food and drink sector has been slow to adopt advanced technologies.[9] Despite being the largest sub-sector in manufacturing in terms of Gross Value Added (GVA),[10] the UK food and drink sector continues to be seen by some in industry as lagging in regard to the take up of industrial digital technologies (IDTs).[11]

This slow response is concerning when digitalisation presents many opportunities for the food and drink sector, including improved planning and forecasting; lower costs; and a boost in productivity, efficiency and profits.[12] Also relevant to the food and drink industry is the need for optimal resilience, regardless of circumstance. The need for innovation is of particular importance to the UK, with the country's high population within a relatively small geographical area being a significant factor in the sector's reliance on supply and demand, as well as global production - both of which can be particularly volatile.[13] This presents a particular need for supply chain models that allow for sufficient visibility and efficiency, which could both be improved by the uptake of IDTs that are commonly associated with the Industry 4.0 movement. Reports have also indicated that the adoption of digital technologies could result in productivity improvements between £7.4 and £11.5 billion for the UK food and beverage sector.[14]

[9] https://ktn-uk.org/news/digital-transformation-of-the-food-and-beverage-sector/
[10] https://www.makeuk.org/insights/reports/manufacturing-outlook-2021-q1
[11] https://www.raconteur.net/technology/why-the-food-and-drink-sector-needs-to-digitalise-now/
[12] https://www.themanufacturer.com/articles/the-food-and-drink-sector-ready-for-its-own-digital-revolution/
[13] https://core.ac.uk/download/pdf/288374238.pdf
[14] SFS, The Digitalization Productivity Bonus, April 2017 and SFS, CFO 4.0, Essential financial competencies for digital transformation in UK manufacturing, April 2018.

We are seeing a shift in the attitude of globalised food and drink companies towards advanced technology solutions, which can play a vital role in filling existing industry gaps and making supply chains more resilient, robust and adaptive. In 2020 Gartner identified[15] that artificial intelligence, edge computing and digital supply chain twins would be among the top trends for supply chain leaders looking to transform and adapt their organisations, while Gartner[16] found that over half (55%) of surveyed supply chain leaders expect to become highly resilient in the next two to three years, demonstrating their appetite to build strong resilience and mitigate risks quickly.

Despite these encouraging figures, uptake remains slow. Reasons include the sector's historical reliance on manual labour - more so than other manufacturing subsectors, the high number of small and medium-sized enterprises (SMEs) that operate in the industry[17] and a number of barriers that industry experts identified during the policy engagement session. These barriers can be grouped into the categories of **knowledge, skills and culture**; **industry relationships**; **technology and innovation landscape**; and **business capacity**.

# Barriers to adoption

## Knowledge, skills and culture

As noted earlier in this report, the food and drink sector is predominantly composed of small and medium-sized enterprises. However, SMEs across the entire manufacturing sector typically have poor rates of adoption compared to their larger counterparts,[18] which may be attributed in part to an understanding around IDTs, what is required to successfully engage with them and what the long-term benefits can be in digitising their processes and supply chains.

**Although lagging in its adoption of advanced digital technologies, in recent years, the food and drink manufacturing industry has started to implement robotics and basic modes of automation to supplement manual labour.** As such, many of the skills currently being used in industry may require further development to upskill employees to the levels necessary to achieve the technical goals within a company. Industry experts have noted that while the intention to upskill employees exists among many manufacturers, including those in food and drink , the quality of training can hinder this, costing valuable monetary and time resources for training that does not meet the required standards.

[15] https://www.gartner.com/smarterwithgartner/gartner-top-8-supply-chain-technology-trends-for-2020/
[16] https://www.gartner.com/en/supply-chain/trends/weathering-the-storm-supply-chain-resilience-in-an-age-of-disruption
[17] https://www.themanufacturer.com/articles/the-food-and-drink-sector-ready-for-its-own-digital-revolution/
[18] https://www.madesmarter.uk/media/y12d3ywe/20171027_madesmarter_final_digital.pdf

In addition to the challenge of upskilling employees is the capability to recruit people with the right skillset, particularly when a company is already short on technical skills. Here, the issue of knowledge is of crucial importance, as those making hiring decisions are required to have a certain degree of knowledge on the subject area to identify what skills and experience are required and how to successfully integrate these skills into the company.

The culture and openness to innovate is also a significant barrier to adoption in industry. As we know, the food and drink industry is largely dominated by manual processes and traditional methods, which is reflected in the fact that the sector has relatively lower levels of R&D intensity than other manufacturing subsectors.[19]

## Industry relationships

UK consumer trust in the food and drink industry is particularly high, thanks largely to the response of the sector in the midst of the COVID-19 pandemic.[20] However, industry experts have raised the concern that this same level of trust is not consistent within the food and drink supply chain.

The inherent nature of a supply chain means that certain levels of good faith, trust and strong professional relationships are a core component of any operations and mitigating risk, even more so when attempting to use digital technologies to improve this mitigation of risk. Session attendees have noted that, in order to successfully adopt industrial digital technologies throughout the supply chain, the elements of trust between parties and increased collaboration, as opposed to pure competitiveness need to be built throughout the chain. While it is true that for profit organisations, component parts of the supply chain are likely to emphasise profitability, cost reduction and efficiency, these parts are mainly working in silos. The most successful supply chains are likely to be ones that are fully collaborative. Research suggests that working collaboratively in a network, as opposed to a linear supply chain, could present an effective organisational structure for digitised supply chains.[21] Through collaboration, this organisational strategy could increase visibility throughout the supply chain, improve communication and reduce the potential for bullwhip effects throughout the supply chain.

## Technology and innovation landscape

A commonly noted concern throughout various sub-sectors of manufacturing, with food and drink being no exception, is the difficulty in ensuring successful interoperability between IDTs, legacy information and technology systems already existing within companies.[22] Industry experts have expressed that difficulties in interoperability affect the successful extraction of data from any automated and digitised equipment that could be useful in reducing inefficiencies and maximising value.

[19] Make UK, Sector Bulletin: Food & Drink
[20] https://www.specialityfoodmagazine.com/news/trust-in-food-industry-at-all-time-high
[21] Digital supply chain: challenges and future directions, Blandine Ageron, Omar Bentahar, Angappa Gunasekaran, 2020 <https://www.tandfonline.com/doi/full/10.1080/16258312.2020.1816361>
[22] KnowRisk Policy Engagement Sessions May, 2021

There is also a perception by some in industry that the available technology is not always affordable or easy to use for most SMEs, with current solutions appearing primarily to cater for larger organisations.[23]

## Business capacity

Session attendees have noted that 96% of the UK's 7,400 food and drink manufacturing businesses are classified as SMEs.[24] However, the nature of being an SME, typically with fewer resources than their larger counterparts, means that many of these businesses are focussed on the here and now with little time to scout for future innovation.

Industry experts consider the existing level of innovation within food and drink SMEs to be a factor in digitalisation. As previously noted, the sector is less technologically mature than many other manufacturing subsets and although the use of both stationary and collaborative robotics is starting to become more common in industry, along with other technologies, the sector as a whole is not highly automated.[25]

Session attendees noted that more widespread automation, with increased entry-level technologies could be a precursor to  digitalisation, as it could begin to demonstrate more efficient processes, upskill sector employees and start the internal operations needed to produce a digital transformation that can trickle down to peers and other SMEs within a supply chain.

[23] KnowRisk Policy Engagement Sessions May, 2021
[24] KnowRisk Policy Engagement Sessions May, 2021
[25] KnowRisk Policy Engagement Sessions May, 2021

# Risks associated with food and drink supply chains

Although the nature of the food and drink industry means that it has typically been resilient to economic depressions,[26] industry experts have revealed several risks that remain within the supply chain that could negatively impact supply chain members.

## Changing trends, perceptions and expectations

This risk includes the demand-driven change in consumer behaviour, such as an increased interest in certain food types and variety of food, which in turn is linked to more inventory and therefore more waste.
Growing changes in UK consumer attitudes towards wellness, including a reduction in meat, sugar and salt consumption,[27] have influenced both policy and food manufacturing guidelines, such as the Soft Drinks Industry Levy (sugar tax) of 2018 which imposed a charge of 24 pence on soft drinks with a specific sugar content. Though celebrated by health and wellness advocates, some food and drink manufacturers were adversely affected, with Coca-Cola noting an impact on sales following the Levy's implementation.[28] Conversely, other trends have proved beneficial for some in industry, with supermarkets noting a rise in sales of meat-free goods and vegan options, as well as the opportunity to diversify their offering with dedicated vegan ranges.[29]

[26] Make UK, Sector Bulletin: Food & Drink
[27] https://www.rand.org/content/dam/rand/pubs/research_reports/RR4300/RR4379/RAND_RR4379.pdf
[28] https://www.beveragedaily.com/Article/2018/10/29/Sugar-tax-knock-for-Coca-Cola-Classic-in-Great-Britain-but-Zero-Sugar-up-50
[29] https://www.bbc.co.uk/news/business-44488051

## Natural disasters, disease and geopolitical concerns

These concerns focus largely on issues surrounding biodiversity and climate change and the ways in which they impact the supply of raw ingredients. Research suggests that food security is becoming increasingly threatened by climate change, with factors such as changing patterns of precipitation and increased temperatures threatening some agricultural production.[30] With an increased number of what should be once-in-a-lifetime events linked to climate change, supply chain partners are faced with an increased risk of unpredictable supply, which is not always fully communicated across the supply chain. When paired with geopolitical concerns including conflicts and trade wars that can significantly affect food security and profitability both within the UK[31] and in developing countries,[32] supply chain partners can be faced with an increased number of unexpected short-term risks that can negatively impact growth.

## Regulatory

Session attendees noted that risks in this category include an increasing number of what are perceived to be litigious stakeholders and interventionist governments, as well as various changes in the regulatory environment associated with the UK leaving the European Union. Clarity over changes in regulation for UK food and drink industry companies, as well as the time required to implement these changes, have been raised as a risk to industry.[33]

## Ethics and financial markets

Issues around the sustainability of food and drink products, as well as the ethical origins of raw goods and finished products have been raised as a key concern in supply chain management. With increased focus from government and consumers on moving towards net zero, many in industry have pledged, and have been making strides, towards sustainable goals such as reducing food and packaging waste, as well as water consumption and $CO_2$ emissions.[34] Long-term, while some of these changes can financially benefit supply chain parties, initial costs and other expenses often require significant financial investment, with costs often coming in higher than initially predicted.[35]

Similarly, issues affecting the financial markets have been shown to have an impact on the purchase of raw materials and goods throughout supply chains, leading to higher prices for consumers. Research suggests that the cost of food has risen at the fastest pace in over a decade.[36]

[30] https://www.ipcc.ch/srccl/chapter/chapter-5/
[31] https://www.foodmanufacture.co.uk/Article/2020/01/29/Union-joins-calls-to-remove-tariffs-affecting-US-and-UK-whiskies
[32] https://unctad.org/news/trade-wars-are-huge-threats-food-security
[33] https://www.pinsentmasons.com/out-law/news/uk-food-drink-companies-struggle-comply-brexit
[34] https://www.leisurefb.co.uk/news/21208/
[35] https://www.bbc.co.uk/news/business-57353624
[36] https://www.bbc.co.uk/news/business-57353624

# Conclusion and recommendations

As the manufacturing sector's largest contributor, with a supply chain structure that can be subject to unpredictable risk, the food and drink industry needs to take advantage of advanced industrial digital technologies.

Through the digitalisation of the supply chain, the sector can reduce risks, increase efficiency and provide an operating environment conducive to maintaining high consumer trust.

Despite the sector's slow uptake of these technologies, several early adopters can demonstrate to the rest of industry the benefits of these technologies in reducing supply chain risk, particularly amongst SME early adopters.

The benefits of collaboration and communication between food and drink supply chain parties is also likely to have a significant impact in creating transformative change within industry, allowing for increased visibility within networks and the continued strengthening of industry.

Feedback on the proposed KnowRisk solution suggested that it could support these goals, improve efficiency and reduce risk by providing an extended visibility of the supply chain, reduce waste and allow for improvements in collaboration. Session attendees also noted that the solution could reach optimal value if it is able to assist with expectation management and opportunities to improve sustainability, as well as scaling the exchange of data.

Insights from the policy engagement sessions conducted have revealed several other potential solutions for reducing risk, noting what is required from a policy perspective as well as required by industry.

## Policy solutions

- Supporting **evidence-based and business literate governments' understanding** of the food and drink challenges.
- Conducting **overall systemic risk assessment** at a national level that feeds the national food and drink sector strategy.
- Establishing **key touch points for a data-driven understanding of the issues** faced that require cross-organisational collaboration.
- Informing **standards** - now and in the future as the ecosystem progresses.

## Industry solutions

- Forming a **collective voice** through the Food and Drink Sector Council.
- Setting a **long-term framework** within businesses, as well as a clear agenda and actions.
- Incorporating some industry-agreed **guiding principles.**
- Identifying areas in which **pre-competitive collaboration** could exist.

Insights obtained from the KnowRisk policy engagement sessions, as well as literature, suggest that sufficient awareness of the blockers preventing digitalisation in food and drink supply chains, as well as the implementation of several of the aforementioned solutions, has the potential long term to enact transformative change in reducing risk.

Focusing on these factors, while also considering known and unknown risks, could help to continue the strength of the food and drink industry, particularly as it looks ahead following several shocks in the past few years.

# KnowRisk: FEDERATED LEARNING AS A SERVICE

KnowRisk:

# FEDERATED LEARNING AS A SERVICE

Addressing some of the key challenges organisations face when adopting a federated learning approach.

**A federated learning approach is essential when one or more data owners need to adopt machine learning solutions that are trained on and run using distributed confidential data.**

This report outlines federated learning as a service (FLaaS), a solution offering to help address some of the key challenges faced by organisations when adopting a federated learning approach to training a machine learning model.

As stated in VentureBeat[1], there are three key challenges when adopting a federated learning (FL) solution: the first is in determining the proper incentive structure so that data owners are willing to participate in the federated training with each other when they may be competitors in the same industry; the second is setting up the correct governance structure so that data is handled appropriately and participants can be confident of that fact; and the third challenge is to understand what additional technical infrastructure is needed (compared to training standard machine learning models) building it, then subsequently developing the model architecture, running the training and deploying the models.

[1] 1https://venturebeat.com/2021/02/12/how-to-know-if-federated-learning-should-be-part-of-your-data-strategy/

# FLaaS: aims and challenges

The primary aim in creating federated learning as a service will be to address the third challenge: providing an easy way for AI solution providers (whether internal or external to the stakeholders) to set up, train, deploy and monitor a federated learning solution, although other approaches to building such a service have previously been demonstrated.[2] In addition, FLaaS will offer consultancy services to help resolve the other two challenges. Specifically, the consultancy aspect of this service will provide guidance on how to properly incentivise a potential data owner to participate in federated learning with other data owners, and also on how to put in place the proper governance structure to ensure that data confidentiality and model fidelity is maintained in a distributed setting.

Incentives are discussed in more detail in the **use cases section**

This report has been created as part of the KnowRisk project, building on the technical federated learning work (including Digital Catapult's federated learning library) completed during the project. Potentially, the offering itself may be fleshed out more and developed further within the KnowRisk project's next iteration, (whatever form it takes) or as part of another collaborative R&D project that has a federated learning component. This fleshing out will determine the specific use cases to be implemented; the functional requirements from the service; the detailed solution architecture and specific technologies adopted; and finally, the implementation and deployment of the service.

The need for this offering is demonstrated with the following set of personas. These personas were used to create a set of use case sketches, using wireframes to illustrate what the user interfaces may look like. As the system being outlined is a fairly significant piece of software, it is beyond the scope of this report to describe the use cases and solution architecture in detail. However, it is anticipated that the open source Digital Catapult federated learning library , developed during the KnowRisk project, will form the engine for federated learning, with the option of users plugging in their own engine should they choose to.

[2] Nicolas Kourtellis, Kleomenis Katevas, and Diego Perino. 2020. FLaaS: Federated Learning as a Service. In Proceedings of the 1st Workshop on Distributed Machine Learning (DistributedML'20). Association for Computing Machinery, New York, NY, USA, 7–13. DOI:https://doi.org/10.1145/3426745.3431337

# KnowRisk FLaaS personas

These personas were created using information obtained during conversations and meetings with members of the KnowRisk consortium and extensive interactions with insurance company personnel in previous projects, prior to KnowRisk. There are two groups of personas: the first group describes three personas from the perspective of AI solution providers; and the second group describes three personas from the perspective of insurance companies.

## AI solution provider

CEO of an AI startup

**Alex is the CEO of a tech startup providing AI-based risk analysis to various businesses.**

She loves going kayaking with her family but has not had many chances to enjoy this during the last year, largely due to her decision to try to enter the insurance market around the same time. The AI models her company builds need data. Alex is having difficulty convincing her potential clients in the insurance industry to release this data, as it is confidential. Additionally, the data is complex and relatively low volume which means that, according to her technical team, they may need to combine data from multiple clients to build a good AI model for the whole industry. This poses both technical challenges in gathering the data and adds to the commercial challenge in convincing clients to not only to release confidential data but to also combine it with data from their competitors. Alex believes that the insurance industry is extremely promising but is beginning to wonder if it's worth the hassle of dealing with their data and the personal cost of lost time with her family. She is actively searching for a solution to this problem and also pushing her technical team hard to find something that will work. If a solution can be found, it will have a massive benefit, opening up other markets where data confidentiality is a barrier to entry.

## CTO of an AI startup

**Corrin is the CTO of a tech company that provides AI-based risk analysis solutions to businesses.**

He is a self-confessed lifelong nerd and is revered among friends for being an imaginative and brilliant Dungeons & Dragons (D&D) dungeon master. However, he wishes his imagination could currently help him a bit more in his job. His company has recently decided to move into the insurance market, a move he has been nervous about.  Unfortunately, his fears have been proven right; the company is having a hard time getting the insurance companies to share their data and furthermore, the complexity and the low volume of the data mean that the data has to be combined to build the AI model. He has heard of federated learning and thinks something like that may just be what is needed. Unfortunately, this is a relatively new piece of technology and this company does not have the resources to experiment with it to adopt existing frameworks or build a solution from the ground up. He is also not sure what the implications are for his tech stack or what kind of model development strategy will be needed in this kind of federated setting. He is happy to leave the incentive and governance aspects to the CEO, but really wishes there was a readily-available tech solution or service for his problem. If so, he could focus on building out the core technology of his company and engage his imagination on more fantastical realms.

## Machine learning engineer

within an AI product company

**Pat is a machine learning engineer at an AI company, which has recently decided to try to target its risk analysis product for the insurance market.**

She has been a ML engineer for about three years and this is her first job after completing her physics undergraduate degree. Pat plans to go back to school and pursue a graduate degree in experimental physics and is confident her machine learning experience will give her an edge. In her spare time Pat enjoys going for bike rides, hiking with her dog and browsing physics journals, but she is currently distracted by the needs of her job, as her boss is asking her to figure out how they can adopt federated learning. Pat understands the general principles and can implement the algorithms she reads in papers but is having trouble translating that into a distributed deployable application because she has not done software engineering of that nature in the past and she does not have the bandwidth and resources to learn, experiment and develop those skills. She wishes there could be a pre-built/low-code or no-code solution to her problem, then she could spend more time indulging her passion for physics.

## Insurance company

Chief Innovation Officer/VP of Innovation
at insurance company

**Ahmed's life and career within the financial industry has taken him all over the world and he has called multiple countries across different continents his home.**

This background gives him an international outlook and he is able to contextualise his innate deep curiosity through that perspective, while understanding the value of tradition. This particular trait makes him perfect as the chief innovation officer at a large multinational insurance firm. Ahmed needs his organisation to embrace the latest and greatest cutting-edge technologies that will address the challenges the business is facing. Culturally, insurance is a conservative industry, so Ahmed needs to make sure that any new technology that is adopted is very likely to succeed and will fit into the existing vision of the technological trajectory of the company. As a result, his main goal is to seek out offerings and services that are not only creating a buzz in the tech industry, but also as reliable as something offered by a big tech firm. Ahmed has met the CEO of a risk analysis firm that uses AI and likes her and what the company is offering, so he really hopes in their upcoming meeting, she will have a solution to the data sharing issue they have discussed. Ahmed is willing to work with her, but if he does not hear anything convincing, he may have to move forward with other less exciting companies. Sometimes he considers moving into the tech sector himself to give his curiosity full reign.

## Head of IT

an insurance company

**Hwang is the head of IT of the Taiwanese branch of a large multinational insurance firm.**

Dealing with IT problems is not her favourite job in the world, so she has worked very hard to move to the top of the organisation where she can focus on making sure the tech stack and services offered are strictly scoped out, properly restricted, exactly what is needed by the employees (nothing more, nothing less) and also highly secure. Her bosses appreciate her approach and this is also perfectly in line with her generally risk-averse personality. She lets her risk-taking side run wild only when she is questing in World of Warcraft, where her gall and nerve is the stuff of legend. However, back in the real world, the recent proposal she received from an AI-based risk analysis company about sharing data between other branches for some project has left her cold. She has a sinking feeling that the VP of Innovation, who is much more adventurous than she would like, will view the proposal approvingly and she plans to fight tooth and nail to prevent the adoption from happening. She likes to keep her real-world kingdom firmly within the realms of reliable, well understood services, for which there is a clear line of responsibility and custody for when things go wrong.

## Head of data science

an insurance company

**For James, both his personal and work time is taken up by technology. He is utterly fascinated by it.**

In his spare time James is a maker, his workspace at his family home is cluttered by raspberry PIs, DIY robots, 3D printers and countless other tech gadgets. At work, where he is the head of data science at a large multinational insurance firm, he deals both with the challenges of doing actual data science work and working within the restrictive tech environment provided by the insurance IT department. His division has his own hardware and network setup for the experimentation needed by his team, but to access the data within the insurance firm, he needs to work under the same stifling constraints as everybody else. In the past, he has been forced to pass on internal projects because his team deemed the restrictions in place too severe for the work involved. James really likes the work being proposed by the new AI-based risk analysis firm the VP of innovation has been talking to, but like other ambitious proposals preceding it, he fears it will be doomed to fail unless it is backed up by some solid and well-thought-out deployment options that fit with the restrictive environment of the insurance firm.

# Use cases

## Introduction to use cases

Due to the large number of use cases it would be beyond the scope of this report to describe them in full. Therefore, the use cases are presented at a high level rather than using standard templates.

Based on the personas, discussions with various partners and the current state of the art, the following approach is recommended: the FLaaS will be offered via an online portal (called the FLaaS portal from now on) that lets a FLaaS user (an AI solution provider) deploy a federated learning solution for their stakeholders (one or more clients/customers of the solution provider).

The FLaaS user may be:

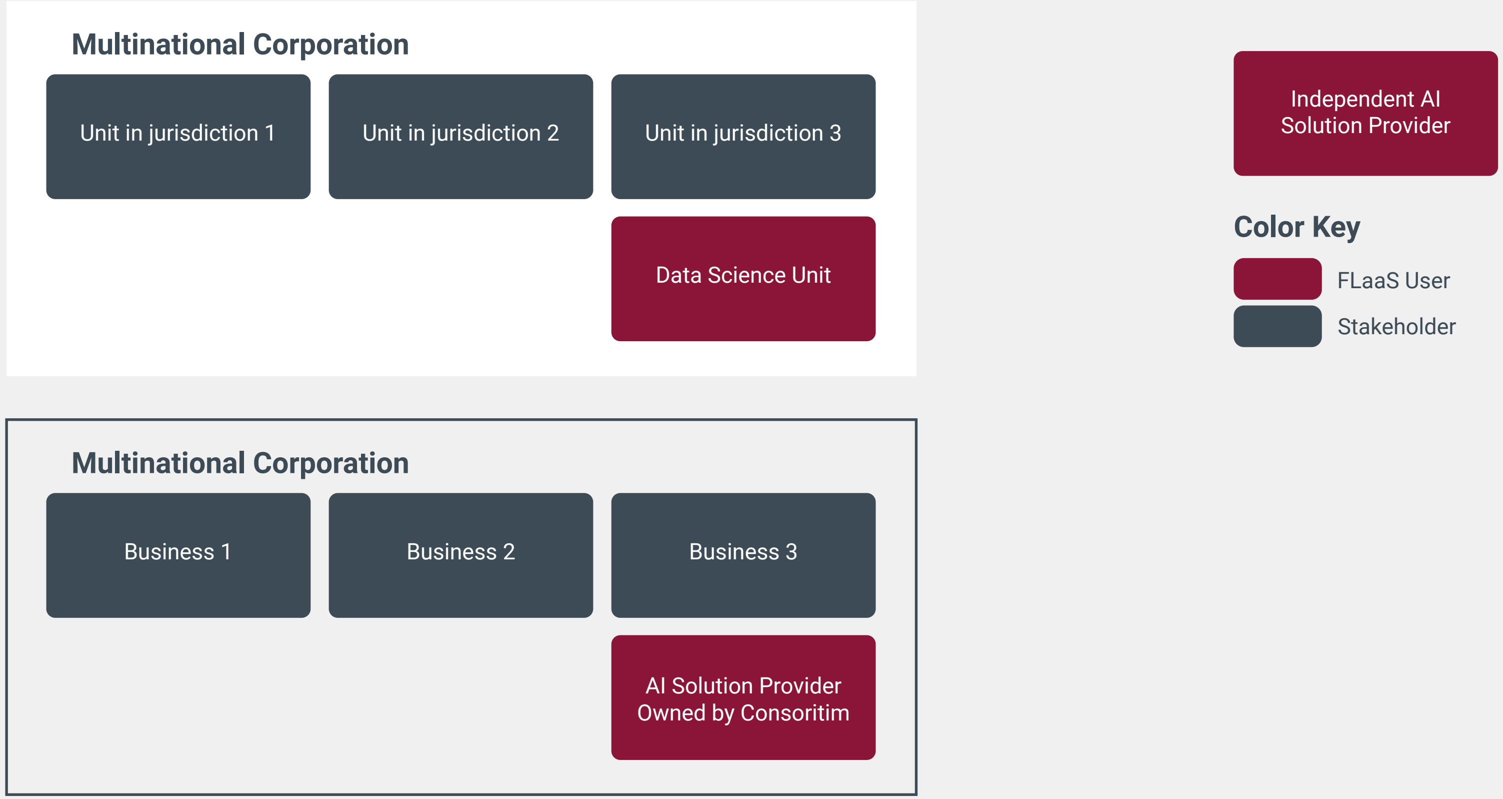
- **a fully independent business**
- **part of some larger enterprise**
- **owned by a consortium of business entities.**

In the context of KnowRisk, a FLaaS user will join as a partner, such as Intelligent AI or Cystellar, who provide AI-enabled services. Following on from that, the stakeholders may be:

- **a single business with jurisdictionally separate units that cannot legally share data (such as a multinational corporation)**
- **a formal consortium of businesses**
- **clients of the FLaaS user in the same industry.**

**Figure 1** illustrates the different possibilities above for the FLaaS user and the stakeholders

In the context of KnowRisk, the stakeholders will be insurance companies and the clients of insurance companies (i.e. businesses) who own confidential data (such as insurance risk reports) that can be used to build an AI-based risk analysis model. This corresponds to the third row in Figure 1.



Multinational Corporation

Unit in jurisdiction 1

Unit in jurisdiction 2

Unit in jurisdiction 3

Data Science Unit

Multinational Corporation

Business 1

Business 2

Business 3

AI Solution Provider Owned by Consoritim

Independent AI Solution Provider

Color Key

🟥 FLaaS User

⬛ Stakeholder

Figure 1: Various possibilities for FLaaS users and stakeholders.

## Use cases sketch

At a high level, the FLaaS portal offers the following services:

- helps the FLaaS user set up the infrastructure necessary for performing federated learning over data owned by the stakeholders
- manages federated analytics to understand the data and subsequently determine the AI model architecture
- manages federated learning sessions on the architectures chosen
- deploys trained local models at stakeholders system for use in production
- offers guidance on incentivising federated learning, plus data and model governance.

The portal will provide a web-based interface that will be used by engineers working for the FLaaS user to set up and deploy the federated learning infrastructure, as well as initiate and manage federated training sessions. In addition to this, the portal will also help engineers and data scientists on the stakeholder side install a FLaaS client service that will operate inside the stakeholder systems, run the worker side logic for federated analytics and deploy and monitor trained local models. **Figure 2** shows this high-level design.

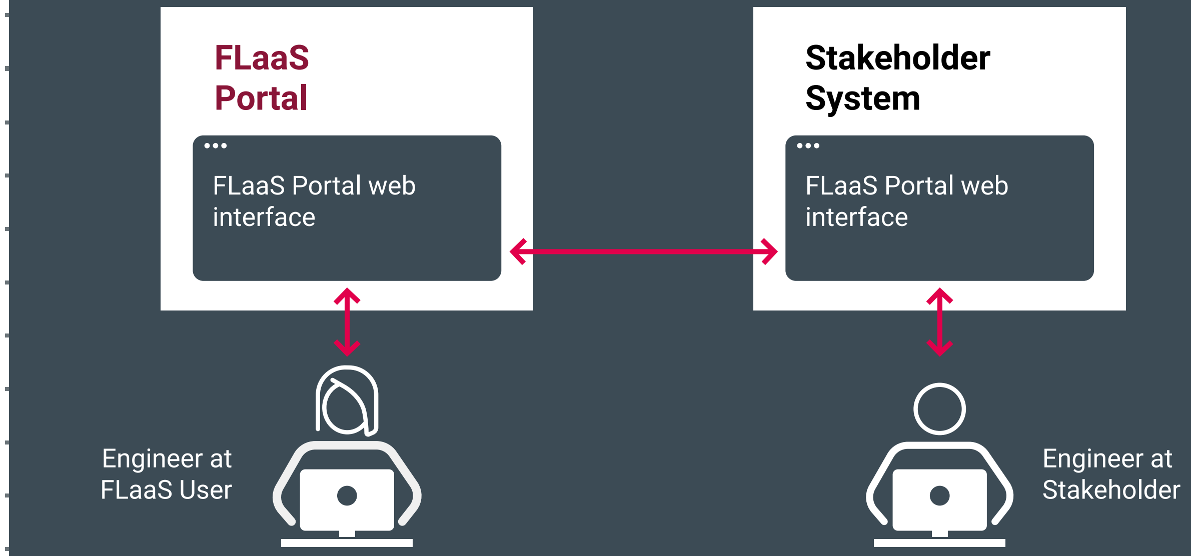


Figure 2. High level FLaaS offering. The figure shows only one stakeholder - however, a component identical to this will be present for each stakeholder.

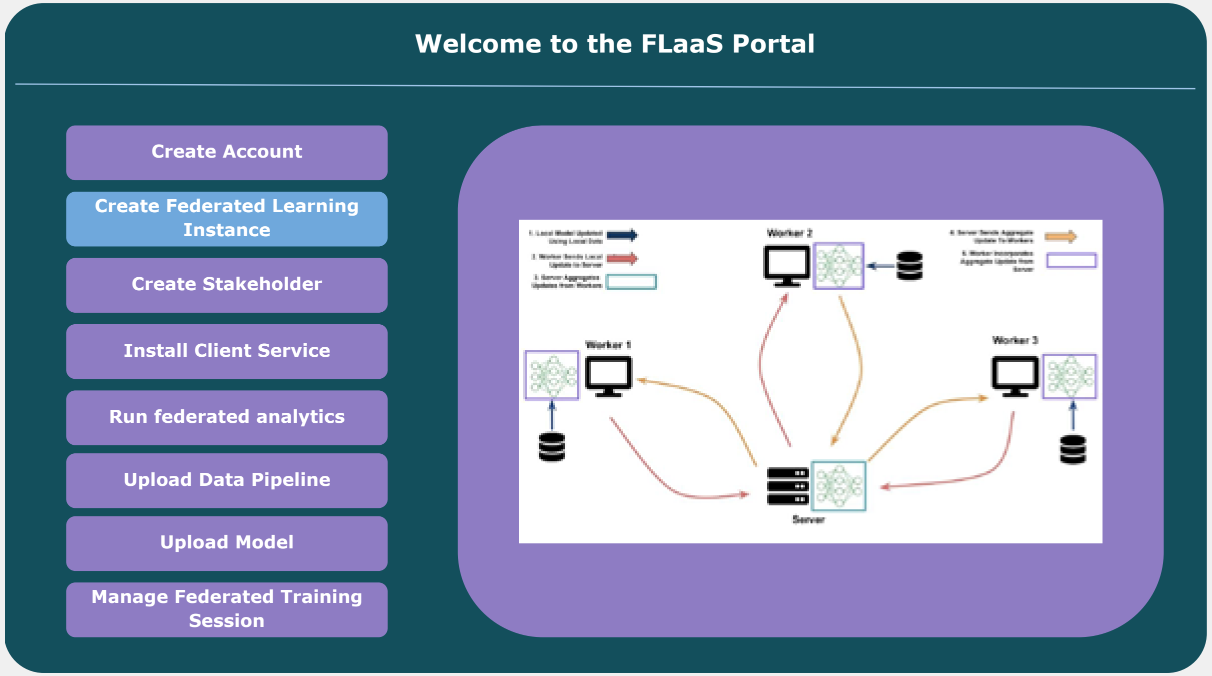

Welcome to the FLaaS Portal
Create Account
Create Federated Learning Instance
Create Stakeholder
Install Client Service
Run federated analytics
Upload Data Pipeline
Upload Model
Manage Federated Training Session
1. Local Model Updated Using Local Data
2. Worker Sends Local Update to Server
3. Server Aggregates Updates from Workers
Worker 2
4. Server Sends Aggregate Update To Workers
5. Worker Incorporates Aggregate Update from Server
Worker 1
Worker 3
Server

Figure 3: Interface for the FLaaS web portal. Each button on the left corresponds to at least one high-level use case. The sequence of buttons also illustrates the overall journey for the user.

## Use cases: FLaaS portal web interface

**Figure 3** shows a wireframe for the web interface of the FLaaS portal. The figure indicates the set of use cases and the journey the user will take when using the portal. The steps in this journey each correspond to at least one complete use case:

1. Create an account with the portal (actor: Engineer at FLaaS user).
2. Create a federated learning instance that will manage all aspects of federated learning for a given group of data owners (actor: Engineer at FlaaS user).
3. Create a stakeholder and possibly add them to a FL instance (actor: Engineer at FlaaS user).
4. Install the client service on the stakeholder systems that will handle activities related to federated learning in a privacy-preserving manner - see below (actor: Engineer at Stakeholder).
5. Run federated analytics to analyse and understand the data in a privacy-preserving manner, so that proper data pipelines can be built for feeding the data into the ML model and then determining the machine learning model architecture (actor: Engineer at FlaaS user).
6. Upload a specific machine learning model architecture for future federated training (actor: Engineer at FlaaS user).
7. Manage and run a federated training session (actor: Engineer at FlaaS user).

## Use cases: FLaaS client service

**Figure 4** below shows a wireframe for the interface of the FLaaS client service. The figure indicates the set of use cases and the journey the user will take when using the portal. The steps in this journey each correspond to at least one complete use case:

1. Connect a data source at the stakeholder side to the FLaaS client so that the data can be consumed by the federated analytics at the FLaaS portal or by the local model via the data pipeline (actor: Engineer at stakeholder).
2. Authorise the FLaaS portal to perform privacy federated analytics on the data (actor: Engineer at stakeholder).
3. Authorise the FLaaS portal to carry out federated training (actor: Engineer at stakeholder).
4. Install a data pipeline created by the FLaaS user engineer, so that the data can be transformed properly to be used by the local ML model (actor: Engineer at stakeholder).
5. Manage the federated training of the local model by choosing the proper data source, setting some model training parameters and selecting update strategies etc. (actor: Engineer at stakeholder).

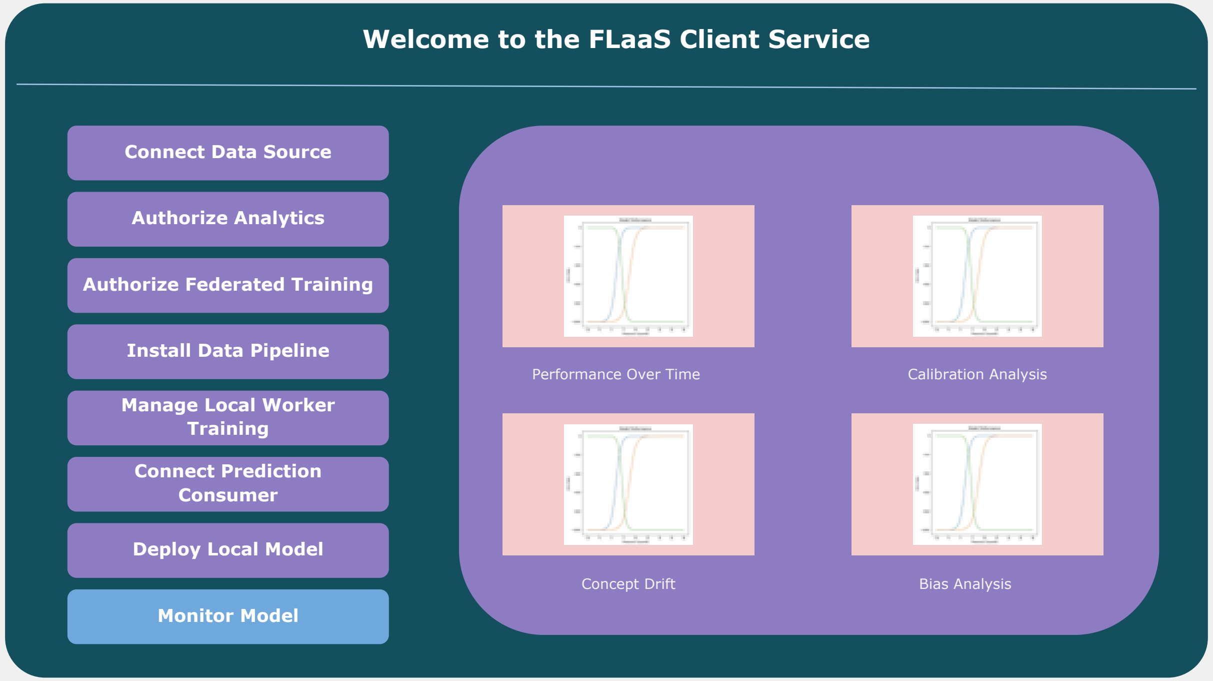


Figure 4: Interface for the FLaaS client service. Each button on the left corresponds to at least one high level use case. The sequence of buttons also illustrates the overall journey for the user.

6. Connect the prediction output of the local model (trained or otherwise) to a prediction consumer, for instance: a dashboard in the AI solutions provider application, some stakeholder risk analysis system (this way the trained model can be used in production for useful tasks - see next step) (actor: Engineer at stakeholder).
7. Deploy the local model so that the model can be used in production, consuming data from a connected data source via an installed data pipeline and sending predictions to a prediction consumer (actor: Engineer at stakeholder).
8. Monitor the local model, which includes but is not limited to: performance tracking, calibration, retraining, change in data format or semantics, error analysis, bias analysis, outlier analysis (actor: Engineer at stakeholder).

## Incentivisation and governance

As mentioned in the introduction, unlike standard machine learning deployment scenarios, **federated learning may require incentivising the data owners to participate in the federated training of models because their data may also benefit their competitors.** The incentive mechanism will typically be bespoke, but the FLaaS portal will also offer consultancy services that will help establish effective incentive mechanisms. Generally speaking, data owners will be unwilling to participate in FL if they perceive that this will cut into their competitive advantage. So, in these cases, FL should target a problem that either reduces cost or improves the quality of a public good in the industry.

As an example, within KnowRisk the data owners are insurance companies who believe their competitive advantage comes from the specific risk categorisations they track and the risk scores they produce. Effective AI-based extraction of some base set of risks and mitigation from PDF risk reports, which then feed into the insurer-specific logic for creating the risk categorisations and risk scores, can be considered a public good that will benefit all insurance companies without affecting their competitive advantage. Thus, extracting risks and mitigations from risk reports using federated learning can be incentivised and the FLaaS consultancy service will help with this work.

.

**Federated learning also requires having proper governance mechanisms in place so that participants can be confident that their data is secure and the model is robust to withstand the influence of data from other participants.** The FLaaS platform will help with this by using standard technologies (cryptographic or otherwise) to ensure that the FLaaS client does not reveal any confidential information via its analytics service and that the models exchanged are also secure. This can be accomplished by adopting technologies such as secure multi-party computation and homomorphic encryption; differential privacy; and testing the models learned to see to what extent (if at all) they can be reverse engineered. Additionally, to ensure that models are not corrupted by bad data or models from other participants, the FLaaS training will incorporate robust federated training mechanisms which will detect and prevent model corruption. Finally, the FLaaS portal may also provide template governance documents to make the process easier.

## Conclusion

This document outlined a federated learning as a service offering through a set of personas, tailored to the KnowRisk project, along with broadly applicable use case sketches.

The goal now is to flesh out the use cases, design the solution architecture and build the offering as part of a commercial project or commercial research and development project. Digital Catapult firmly believes that such an offering will be necessary to fully realise the potential and benefits of federated learning in ensuring powerful and beneficial AI solutions are widely adopted.

# About Digital Catapult

Digital Catapult is the UK authority on advanced digital technology.

Through collaboration and innovation, we accelerate industry adoption to drive growth and opportunity across the economy. We bring together an expert and enterprising community of researchers, startups, scaleups and industry leaders to discover new ways to solve the big challenges limiting the UK's future potential.

Through our specialist programmes and experimental facilities, we make sure that innovation thrives and the right solutions make it to the real world. Our goal is to accelerate new possibilities in everything we do and for every business we partner with the journey — breaking down barriers, de-risking innovation, opening up markets and responsibly shaping the products, services and experiences of the future.

Digital Catapult is part of the Catapult Network that supports businesses in transforming great ideas into valuable products and services. We are a network of world-leading technology and innovation centres established by Innovate UK.

Visit www.digicatapult.org.uk for more information.

# Partners

The partner companies involved in the KnowRisk project are SweetBridge, Engine B, Cystellar, Digital Catapult, Industria Technology and Intelligent AI, with Sweetbridge being the leading partner.

## Sweetbridge

is a synchronised commerce platform built on distributed ledger technology, converting any commercial relationship, supply chain or value chain into an ecosystem that increases the net worth of its members.

www.sweetbridge.com

## Engine B

is a digital technology company that combats a major problem for Professional Services firms everywhere - access to quality data. Backed by industry, Engine B's collaborative approach to the development of technology aims to increase openness in the marketplace and create ground-breaking intellectual property for the sector.

www.engineb.com

## Cystellar

is a geospatial intelligence company on a mission to deliver real-time insights for the insurance, logistics and agricultural sector to support data-driven risk assessment and decision making.

www.cystellar.com

## Industria Technology (INDUSTRIA)

is a global technology consulting, ventures and development firm. Industria implements cutting-edge technologies, such as enterprise blockchain, confidential computing, process automation and digital experience to give organisations a clear path to improve performance.

www.industria.tech

## Intelligent AI

is an award-winning AI and data science company focused on the commercial property insurance sector. For insurers, brokers and customers, Intelligent AI provides enhanced understanding of risk, better decision making and improved client journeys through exceptional data insight and real-time document processing using AI, satellite image analysis, data analytics, online risk survey tools and Digital Twins.

www.intelligentai.co.uk