Establishing a social licence for Financial Technology: Reflections on the role of the private sector in pursuing ethical data practices

Mhairi Aitken1, Ehsan Toreini2, Peter Carmichael2, Kovila Coopamootoo2, Karen Elliott1 and Aad van Moorsel2

Abstract
Current attention directed at ethical dimensions of data and Artificial Intelligence have led to increasing recognition of the need to secure and maintain public support for uses (and reuses) of people’s data. This is essential to establish a “Social Licence” for current and future practices. The notion of a “Social Licence” recognises that there can be meaningful differences between what is legally permissible and what is socially acceptable. Establishing a Social Licence entails public engagement to build relationships of trust and ensure that practices align with public values. While the concept of the Social Licence is well-established in other sectors – notably in relation to extractive industries – it has only very recently begun to be discussed in relation to digital innovation and data-intensive industries. This article therefore draws on existing literature relating to the Social Licence in extractive industries to explore the potential approaches needed to establish a Social Licence for emerging data-intensive industries. Additionally, it draws on well-established literature relating to trust (from psychology and organisational science) to examine the relevance of trust, and trustworthiness, for emerging practices in data-intensive industries. In doing so the article considers the extent to which pursuing a Social Licence might complement regulation and inform codes of practice to place ethical and social considerations at the heart of industry practice. We focus on one key industry: Financial Technology. We demonstrate the importance of combining technical and social approaches to address ethical challenges in data-intensive innovation (particularly relating to Artificial Intelligence) and to establish relationships of trust to underpin a Social Licence for Financial Technology. Such approaches are needed across all areas and industries of data-intensive innovation to complement regulation and inform the development of ethical codes of practice. This is important to underpin culture change and to move beyond rhetorical commitments to develop best practice putting ethics at the heart of innovation.

Keywords
Financial Technology, data, social licence, ethics, responsible artificial intelligence, trust

Introduction
Recent years have witnessed a dramatic increase in attention directed at ethical dimensions of data practices and Artificial Intelligence (AI). Increasingly, momentum for innovation is being met with interest in related ethical considerations and a number of high profile institutes and bodies have been established to focus on this area. These include the European Commission’s High-Level Expert Group on Artificial Intelligence whose mandate included drafting a set of AI Ethics Guidelines (European Commission, 2019); the UK Government’s Centre for Data Ethics and Innovation; a House of Lords Select Committee on Artificial Intelligence which proposed an ethical
framework for AI; the Ada Lovelace Institute and; DeepMind Ethics and Society. Internationally, the World Economic Forum has proposed nine ethical questions to ask of AI systems and the Fairness, Accountability, and Transparency in Machine Learning forum developed Principles for Accountable Algorithms and a Social Impact Statement for Algorithms (Stahl and Wright, 2018). Such bodies focus on developing Responsible AI (e.g. PWC, 2018) and trustworthy approaches to AI and digital innovation. While questions remain as to whether, in practice, this goes further than “ethics washing” (Hasselbalch, 2019), this has led to a proliferation of guidance and principles relating to ethical AI (Fjeld et al., 2019) and within the private sector there is emerging interest in the concept of Corporate Digital Responsibility (CDR; Lobschat et al., 2020).

This trend is in part a response to high profile public controversies around data misuse as well as the introduction of new regulation through the EU General Data Protection Regulation (GDPR), which gives individuals greater control over their own data (Politou et al., 2018). In the wake of such developments questions have arisen as to whether compliance with regulation is sufficient to ensure ethical data practices and to what extent ethical codes of practice are also needed (Hasselbalch, 2019). This has resulted in increased recognition of the need to secure and maintain public support for uses (and reuses) of people’s data in order to establish a “Social Licence” (SL) for current and future practices.

The notion of a “Social Licence” recognises that there can be meaningful differences between what is legally permissible and what is socially acceptable (Carter et al., 2015). A SL is granted by a community of stakeholders and is intangible and unwritten but may be essential for the sustainability and legitimacy of particular practices or industries. Developing and maintaining a SL requires public engagement incorporating diverse perspectives and interests, beyond those of professional communities, to ensure that current and future practices are aligned with the values of society.

While the concept of the “Social Licence” is well-established in other sectors it has only very recently begun to be discussed in relation to digital innovation and data-intensive industries. This article therefore draws on existing literature relating to the SL in extractive industries to explore the potential approaches needed to establish a SL for emerging data-intensive industries. Additionally, it draws on well-established literature relating to trust (from psychology and organisational science) to examine the relevance of trust, and trustworthiness, for emerging data practices in the private sector. In doing so the article considers the extent to which pursuing a SL might complement regulation and inform codes of practice to place ethical and social considerations at the heart of industry practice.

In order to illustrate this, we focus on one key industry: Financial Technology (FinTech). We argue that as a data-intensive industry FinTech requires ethical data practices to be developed and demonstrated in order to establish and maintain a SL. While there is industrial advocacy surrounding the potential benefits of data science and AI in banking, it is not yet clear whether there is a SL for these practices. Therefore, FinTech provides a timely example through which to examine the opportunities and potential approaches to develop ethical data practices beyond compliance with regulation.

The article draws together the multi-disciplinary perspectives of the authors to reflect on how a SL for data-intensive industries might be realised. It begins by providing some background to the concept of the SL before discussing the ways in which this has been applied to innovation in data practices. The article will then focus on FinTech and discuss the relevance and implications of a SL for the FinTech industry. In particular, the article draws on literature from computer science, organisational science and science and technology studies to consider the importance of developing relationships of trust and to set out some of the technical and social approaches used to facilitate this.

Notes on terminology

Data-driven industry is becoming an established and commonly used term to describe industries whose operations are underpinned by data practices (e.g. data collection, storage, linkage, analysis or sharing), this term implies that the data that is used pre-exists the operations of the industry and that its existence (in an objective and “real” sense) shapes the operations and practices of industry. However, this positivist position overlooks the ways in which data are created through the operations and practices of industry. Indeed the creation and curation of data are equally important functions of these industries, and it is through these functions that data comes to have value (Sadowski, 2019). Therefore throughout this article we refer to data dependent or data-intensive industries. The term “data dependant” acknowledges that access to data and the continuous creation of new data is fundamental to the operations of these industries. “Data-intensive” acknowledges the central role of data in the operations and practices of these industries (both creating and using data).

The use of the term AI emphasizes the algorithms that operate on the data as well as the associated automated or autonomous decision-making carried out by a computer. The ethical concerns in AI extend beyond
companies have referred to the SL in pursuing
activities. Indeed, previous studies have found that
SL can be employed ‘opportunistically to serve the particular objectives and goals
of companies, activists and governments’ (Moffat et al., 2016: 480). This led to criticisms that SL can be employed ‘opportunistically to serve the particular objectives and goals
of companies, activists and governments’ (Moffat et al., 2016: 480). Importantly, relationships
between industry and stakeholders are not static, but rather are continuously evolving and adapting. Therefore, the nature of the SL will also evolve and adapt and ‘business must have regard for evolving social attitudes and expectations if it is to maintain its “social licence’” (Brown and Fraser, 2006: 108).

A social licence for data practices
In response to high profile public controversies around data use (and misuse), recent years have brought calls to establish a SL for data practices (e.g. Allen et al., 2019, Carter et al., 2015; Lawler et al., 2018; Leonard, 2018). For example, the New Zealand Government established the Data Futures Partnership to develop guidelines that public and private organisations can use to develop a SL for data use (Data Futures Partnership, 2017). The resulting guidelines ‘aim to enable organisations to maximise the value of data through building the trust of their clients and developing wider community acceptance’ (Data Futures Partnership, 2017). They consider a SL to be established ‘when people trust that their data will be used as they have agreed, and accept that enough value will be created’ (Data Futures Partnership, 2017).
have discussed the role of data custodians in establishing and maintaining a SL for the use of personal information in health research.

In applying the concept of SL to emerging data-intensive industries comparisons are invited between these new industries and extractive industries in which SL has previously been applied. Data is often depicted as the “new oil” or a “goldmine” (e.g. CTO, 2019; The Economist, 2017): such analogies conceptualise data as a commodity and a resource to be exploited. However, this analogy has been contested as it is noted that data does not exist in a natural form ready to be extracted, but rather it is created or manufactured and given value through the ways in which it is used (Sadowski, 2019). Yet while data may not be the new oil, data dependant industries nevertheless have much to learn from previous (good and bad) experiences of extractive industry engagement with stakeholders to address practical, ethical and social dimensions of their operations. In particular, processes of data accumulation through the creation of new markets and services in previously under-served regions of the world creating new forms of dependence and leading to “data colonialism” (Sadowski, 2019) are reminiscent of the history of exploitative relationships in extractive industries where natural resources have been extracted and removed from local areas to profit overseas corporations. The concept of SL was established as a means of addressing such injustices and as such may play an important role in ensuring data justice (Taylor, 2017).

This brief history of SL in extractive industries highlights the importance of building and maintaining relationships with stakeholders to maintain stakeholder support. The ways in which stakeholders are defined and identified are crucial considerations in pursuing a SL. Typically extractive industries have sought to engage with local communities (as “communities of place”: Moffat et al., 2016) whereas approaches based on physical proximity are likely to be irrelevant to the activities of most data-intensive industries such as FinTech. Narrow definitions might conceive stakeholders as being customers, investors or regulators, whereas broader approaches would include all people from whom data is derived and used in developing or implementing data-dependent services and/or anyone who is potentially affected by the industry’s activities. While GDPR gives individuals greater control over their data, in many instances individuals do not explicitly consent to their data being used in developing or implementing new services (e.g. where data is used in aggregate form), or know how their data might be used. Moreover, people increasingly have limited choices regarding whether or not to use a service which requires access to their data. Given that data practices are having far-reaching – and often unpredictable – impacts across society a broad conception of stakeholders acknowledges the importance of wide public engagement beyond potential service-users. As such wide public engagement with broad publics is vital to ensure that current and future practices reflect public values and interests and has an important role to play in strengthening wider science–society relations (Aitken et al., 2019). The extent to which such broad approaches are likely to be adopted by industry is considered further below.

It is noteworthy that while the concept of the SL originates in controversies around private sector activities, to date, the SL for data practices is largely being examined in relation to public sector uses/reuses of data (e.g. relating to health data). Important questions therefore remain relating to the role of private companies in establishing a SL for current and future data practices. As technologies are increasingly developed and deployed in the private sector, and ethical considerations arise in their application, greater consideration should be given to the role of private companies in addressing these issues. We aim to address this gap in the literature by examining the ways in which a SL might be established in emerging data-intensive industries, using FinTech as an illustrative case study.

FinTech

FinTech has been defined as ‘a new financial industry that applies technology to improve financial activities’ (Schueffel, 2016: 45) and FinTech firms have been described as ‘firms that are combining innovative business models and technology to enable, enhance and disrupt financial services’ (Gulamhuseinwala et al., 2015: 4). Through innovation in FinTech, the financial industry as a whole is evolving and adopting new technologies and data-intensive practices. Innovations which have played a role in creating this impact include internet banking, mobile payments, crowdfunding, peer-to-peer lending, Robo-Advisory and online identification (Schueffel, 2016). Some of the technologies used by FinTechs include: User-facing web-based technologies including applications for mobile phones and web browsers; back-end technologies, such as cloud and blockchain; data collection, processing and analytics. AI in particular is used for a range of purposes including developing automated chatbots for customer services, efficient processes for detecting fraud and money laundering and improving automated processes that utilise large volumes of data (e.g. client risk profiling or credit scoring; Maskey, 2018).

Customer uptake of FinTech products is rapidly expanding: In 2015, the EY FinTech Adoption Index reported that ‘a weighted average of 15.5% of digitally
active consumers are FinTech users’ according to their definition of a FinTech user using at least two FinTech products (e.g. mobile payments, apps or insurance telematics; Gulamhuseinwala et al., 2015: 6). Just two years later this number had doubled with 33% of survey respondents being FinTech users (Gulamhuseinwala et al., 2017). Thus, FinTech represents a fast-developing industry underpinned by data-dependent technologies. Given that finance is an area that affects most – if not all – members of society the potential impacts of this industry are significant. Such impacts might include transforming the way people access and use money, creating cashless societies (Teigland et al., 2018), opening up financial services to unbanked or underbanked populations (World Bank, 2017) and enhancing competition in the market (Bank of England, 2019). However, simultaneously the reliance on data-intensive technologies and processes can risk creating new opaque systems through which access to finance is determined or to increasingly necessitate citizens’ participation in the Big Data society. As such, the emergence of FinTech represents a timely opportunity to examine the relevance of SL for informing ethical data practices in private sector organisations within data-intensive industries.

To date literature around FinTech has not explicitly engaged with the concept of a SL. Studies which have examined public attitudes or responses have typically focused on customer uptake of FinTech products (e.g. Chuang et al., 2016; Gulamhuseinwala et al., 2015, 2017). In doing so they have tended to focus on customers’ motivations for using FinTech services, and largely neglected non-customers’ reasons for not using FinTech services, or the reasons why some FinTech offerings have been unsuccessful (Kavuri and Milne, 2019). There is a lack of public deliberation or engagement to examine the extent to which FinTech practices align with public values and interests. In short, wider issues around public acceptability of FinTech or its role in society remain largely overlooked.

There are a number of reasons why FinTechs might be motivated to pursue a SL. These reasons in turn reflect different underpinning rationales which can be normative, instrumental and/or substantive (Fiorino, 1990; Wilsdon and Willis, 2004). First, a normative rationale leads to moral positions that suggest that FinTech firms should engage with stakeholders and reflect public values as ‘it’s the right thing to do’ (Wilsdon and Willis, 2004). Second, more practically minded approaches follow instrumental rationales which view efforts to establish and demonstrate a SL as means to achieve an organisation’s own objectives (Wilsdon and Willis, 2004). Instrumental rationales might lead to a variety of potential approaches including: Efforts to build and maintain public trust in order to attract and retain customers; Adopting ethical and transparent approaches to business in order to prepare FinTechs to anticipate and respond to regulatory and policy developments or; Efforts to demonstrate a SL in order to set FinTechs apart from traditional banking institutions. However, following a purely instrumental rationale can lead to approaches which pay ‘lip service’ to public concerns through enacting purely cosmetic forms of public engagement without genuine intentions to address concerns or reflect public values in a company’s operation.

A final set of motivations are underpinned by substantive rationales that regard the development and maintenance of a SL as being aimed at creating wider positive outcomes across society ‘from this point of view, citizens are seen as subjects, not objects, of the process. They work actively to shape decisions, rather than having their views canvassed by other actors to inform decisions that are then taken’ (Wilsdon and Willis, 2004: 39). Following this approach engaging with public values and interests aims to establish a SL for current and future practices while also offering opportunities to “do things better” and maximise benefits not only for the FinTechs concerned but also for wider society. Here it is important to emphasise that establishing a SL is not simply about avoiding or mitigating potential negative impacts but equally about maximising the benefits of FinTech.

In the following sections, we consider ways through which FinTechs could pursue a SL, focussing on the importance of developing relationships of trust and the role of public engagement in establishing and maintaining a SL.

**Trust and trustworthiness in FinTech**

Previous studies have considered consumer trust in relation to FinTech. For example, Gulamhuseinwala et al. (2015) described that while many potential customers look positively at FinTech offerings, more than 25% preferred traditional providers and another 11% do not trust new FinTech companies. Similarly, Chuang et al. (2016) concluded that brand and service trust significantly affect attitudes and willingness to use FinTech services. However, this attention to consumer trust has often overlooked considerations of what it means for a FinTech to be trustworthy. While trust is at the heart of a SL, this is established through mutual relationships enabling all interests and perspectives to be reflected and addressed (Moffat et al., 2016). Customers – and the wider public – do not passively receive information about technologies, products or services (Wynne, 2006). In emphasising the importance of aligning with public values, the SL requires dialogue and public engagement to identify and address public
values, concerns and interests (Moffat et al., 2016). Therefore, establishing a SL requires reflection not just on ways of building public trust but also of establishing and demonstrating trustworthiness (Aitken et al., 2016a, 2019).

**Establishing trustworthiness**

A number of authors have proposed frameworks to examine perceived trustworthiness. For example, Butler and Cantrell (1984) suggested that trust was based on perceptions of: integrity; competence; consistency; loyalty and; openness. Butler (1991) later expanded this list to include: discreetness; fairness; promise fulfilment; availability; receptivity and; overall trustworthiness. Mayer et al. (1995) suggested that perceived trustworthiness is based on judgements of an entity’s: Ability, Benevolence and Integrity (ABI). This ABI framework has since been widely used and further developed to examine relationships of trust in organisational settings.

Ability refers to the extent to which the entity is perceived to have the skills and competencies to carry out the particular tasks relevant to the situation in which they would be trusted. Benevolence is described as ‘the extent to which a trustee is believed to want to do good to the trustor’ (Mayer et al., 1995: 718). Integrity requires confidence that the entity will act in accordance with a set of principles and that those principles align with the values of the trustor (Mayer et al., 1995). If an entity is perceived to possess each of these attributes they are likely to be trusted, whereas if ‘any of these attributes [are] called seriously into question, this makes us wary’ (Dietz and Gillespie, 2012: 6). Each of the attributes are related and may reinforce one another; however, it is also possible for someone to trust an entity when one or more of the attributes is considered to be lacking as ‘each of the three factors can vary along a continuum’ (Mayer et al., 1995: 721).

More recently authors have expanded on Mayer et al.’s framework suggesting that including Predictability or Reliability is important. For example, Dietz and Den Hartog (2006) suggested that the four key characteristics on which judgements of trustworthiness are based are: Ability; Benevolence; Integrity; and Predictability (referred to as the ABI+ model). Predictability will reinforce perceptions of the Ability; Benevolence; and Integrity of the trustee.

Considering the role of Predictability or Reliability draws attention to the importance of trust being sustained overtime through ongoing relationships. Importantly these relationships should be able to adapt to changing contexts and social dynamics (Moffat et al., 2016). Studies which have examined the conditions needed to establish and maintain a SL suggest Predictability is crucial in relation to openness, transparency and commitments to meaningful engagement (Moffat and Zhang, 2014). Moffat and Zhang (2014) found that procedural fairness and good quality engagement were key to sustaining relationships of trust underpinning a SL. This suggests that predictability in terms of an organisation’s approaches and fairness may be more important than predictability of particular actions or activities which should be able to adapt in response to engagement processes (Moffat et al., 2016). Therefore, in pursuing a SL, FinTechs need to ensure consistency in engagement approaches.

**Relationships of trust**

Relationships of trust can take many forms. For example, there are direct relationships between a FinTech firm and its customers (or other stakeholders), but there are also networks of indirect relationships through which assessments of trustworthiness are made. Andras et al. (2018) note that trust can be developed through awareness of others’ experiences or interactions with the trustee, if someone we know (and trust) uses a particular service we assume that it is trustworthy ‘the main idea, however, is that we did not generate trust in the [service] per se but trust a person that trusts the [service]’ (Andras et al., 2018: 6). Moreover, trust behaviours can reflect trust in a third party whose association with the trustee gives the trustor confidence to take such behaviours – this, Andras et al. (2018) describe as Second Order Trust. For example, when making purchases online, a shopper trusts the online review system (and the anonymous reviewers), which enables them to have confidence in the product or service they are buying and gives them Second Order Trust in the seller. Similarly, Second Order Trust may be established based on trust in financial regulators whose approval may be perceived to give FinTechs “legitimacy”.

The concept of Second Order Trust is particularly salient around new technologies and services, where early adopters are likely to have either high technical knowledge or propensity to risk-taking. Wider adoption of the technology or service will depend on trust building up through social networks emanating from these early adopters. Therefore, while early adopters’ experiences may depend more on confidence in technical competencies and first-order trust in new services, subsequent adopters’ relationships with those services are likely to be founded on Second Order Trust.

Second Order Trust draws attention to the importance of multiple relationships and the ways in which individuals assess the trustworthiness of an entity – such as a FinTech firm. These relationships are both direct and indirect and never static, rather ‘the level of
trust will evolve as the parties interact’ (Mayer et al., 1995: 727). This highlights the relevance of perceived predictability as a factor influencing assessments of trustworthiness (Dietz and Den Hartog, 2006). Dietz and Den Hartog (2006) note that regularity of behaviour over time will strengthen trust whereas unpredictability or unreliability will weaken trust. Furthermore, trust can be either strengthened or weakened through interactions between trustors and trustees as well as through indirect relationships in social networks. As these relationships evolve assessments of an entity’s ABI will also adapt.

**Trustworthy technology**

Given the importance of innovation in data practices for FinTech, a SL for FinTech operations will depend on perceived trustworthiness not only of FinTech firms but also of the technologies underpinning new financial services. As new market entrants, the issue of trustworthiness presents a pivotal challenge for FinTechs where most are yet to establish strong brand reputations. Moreover, both the financial sector and data dependant technologies (such as AI) have been the subject of public controversies in recent years. The public image of the financial industry is still recovering from the effects of financial crash and mortgage crisis of 2008 (Dietz and Gillespie, 2012) and recent years have brought considerable press coverage of scandals relating to mishandling, misuse or abuse of data. Since perceptions of an organisation’s trustworthiness are shaped by context and awareness of related events these factors will be significant in influencing public perceptions of FinTech. Given increasing attention directed at social and ethical dimensions of new data dependant technologies, FinTechs – whose services rely on these technologies – will need to anticipate and address the challenges this presents.

In developing and implementing new financial services underpinned by data dependant technologies a variety of practical and ethical challenges are encountered. Practical considerations include developing mechanisms to ensure security of logins when using banking applications on mobile phones and; minimising risks of privacy breaches. Ethical considerations include ensuring fairness in algorithmic decision-making; avoiding unjust outcomes; ensuring equal access across society to the benefits of technology; considering the potential impacts of automation on perceived responsibility for outcomes (on the part of both professionals and customers) and ensuring that automated processes do not reduce customer autonomy (Scott, 2017). Simultaneously, there are wider ethical issues around data and AI that are of relevance to FinTech, these include concerns regarding surveillance; bias in data and; risks of unemployment through increased automation (e.g. O’Neil, 2016; Stahl and Wright, 2018).

**Technical approaches**

While developing trustworthy data practices remains an emerging field of interest, increasingly technical approaches are being developed with this aim (Toreini et al., 2019). For example, IBM (n.d.) has set out approaches towards ‘building and enabling AI solutions people can trust’ through four key features of “Trustworthy AI”: Robustness, Fairness, Explainability and Lineage. The ways in which these four key features might establish trustworthiness are summarised below:

**Robustness**

The European Commission (2019: 16) notes that technical robustness is a crucial component of trustworthy AI and states that ‘technical robustness requires that AI systems be developed with a preventative approach to risks and in a manner such that they reliably behave as intended while minimising unintentional and unexpected harm, and preventing unacceptable harm.’ Across literatures there are varying definitions of Robustness. Robustness refers to ‘dependability of a system with respect to external faults, which characterizes a system reaction to a specific class of faults’ (Avizienis et al., 2004: 23). Robustness then is a state in which an algorithm functions normally in the presence of accidental faults or a malicious intruder (attacker), while the attacker actively or passively manipulates the operation of the algorithm. In literature around machine learning (Bhagoji et al., 2018; Rauber et al., 2017), robustness includes security but also privacy issues of the algorithm as well as likely barriers to its performance (including errors caused by implementation faults or the algorithm’s accuracy limitations). We use the term robustness in such a general sense.

The bottom line for all defensive approaches is the need for realistic analysis of the potential attackers for goal, knowledge, capability and strategy. Security and Privacy aspects of a machine learning based system have two aspects: safe data and safe model (Liu et al., 2018). The first focuses on the security and privacy issues of the data which is vulnerable against different attacks, more importantly the injection of invalid/malicious input from adversaries or leakage of sensitive information. The second, on the other hand, resolves the security and privacy concerns of the model in terms of reliable functioning and trustworthy performance.
Mathematical modelling of the behaviour of the system is undertaken to identify and mitigate the unpredictable causes of faults in machine learning performance (either due to security issues or implementation and accuracy errors). In these approaches (Hein and Andriushchenko, 2017; Raghunathan et al., 2018), the system is modelled mathematically and its behaviour is analysed in different situations. Such approaches aim to guarantee the predictability of the machine learning system and its resilience to different categories of faults.

Technical approaches aimed at ensuring robustness provide a diverse range of methods to avoid disclosing users’ privacy, maintain the functioning integrity of the AI and remain resistant against attack. As such, these approaches aim to demonstrate Ability through their technical competence to safeguard data, while also demonstrating the organisation’s Benevolence and Integrity in taking measures to protect individuals’ privacy. We posit that demonstrating such characteristics consistently over time will also enhance perceived Predictability or Reliability.

Fairness

Avoiding unfair bias in algorithmic decision-making is a crucial element of trustworthiness. This is particularly relevant for FinTechs that rely on AI to improve efficiency and accuracy in decision-making processes. The European Commission (2019) notes that unfair bias can arise through the inclusion of inadvertent historic bias, incomplete data or a lack of good governance models. If such bias persists in the algorithm it can ‘lead to unintended (in)direct prejudice and discrimination against certain groups or people, potentially exacerbating prejudice and marginalisation’ (European Commission, 2019: 18).

The baseline assumption in fairness-based approaches is that data is biased and should be moderated. Fairness can be addressed in one of the following stages in a model’s operation cycle: pre-processing, algorithm modification and post-processing (d’Alessandro et al., 2017; Friedler et al., 2019). Pre-processing fairness resolutions are focused on the mitigation of bias in the data itself. Resolutions tend to be independent of the AI model. They either re-label the data samples to make the results fair (Jiang and Nachum 2019) or assign a weight to each one where samples that are more likely to be discriminated against receive more attention (Calmon et al., 2017; Kamiran and Calders, 2012). Algorithm modification methods aim to propose AI models that are inherently fair. Such fairness is fulfilled in either a model that is designed to be statistically fair (Kamishima et al., 2012), or through deployment of an auditor that enforces fairness to the model when it is processing the results (Agarwal et al., 2018; Zhang et al., 2018). Post-processing solutions detect discrimination in the outcome of the algorithm. While there are numerous definitions of fairness in the literature, these approaches strongly rely on a mathematical definition of fairness. These solutions measure the fairness of an algorithm by assessing the disparity between privileged and unprivileged groups in the algorithm results. Kusner et al. (2017) categorised post-processing solutions into four groups. In each the data contains one or more protected features that identify the privileged and unprivileged groups (e.g. gender or ethnicity).

1. Fairness through Unawareness discards protected features in the decision-making process. However, this is not a robust solution because it does not consider the correlation between protected features and other features of data (Chen et al., 2019).
2. Individual Fairness considers an algorithm fair if it gets similar predictions for similar individuals.
3. Demographic Parity is satisfied if the prediction results of a group would be the same with or without considering protected features.
4. Equality of Opportunity requires the accuracy of an algorithm to be equal between privileged and unprivileged samples.

There is no comprehensive solution to eliminate discrimination. Therefore, the choice of the technical approaches addressing fairness aim to detect or prevent bias in the output of an AI model. As such they aim to demonstrate Ability through technical competence, Benevolence through avoiding harm to minority or vulnerable groups and Integrity through taking approaches that reflect the values of society (which is pivotal to establishing a SL).

Explainability

AI algorithms are often considered “black box” models (Michie et al., 1994); thus the processes through which outputs are derived lack transparency. The European Commission (2019: 18) states that ‘whenever an AI system has a significant impact on people’s lives, it should be possible to demand a suitable explanation of the AI system’s decision-making process.’ Moreover, the right to an explanation is a key feature of GDPR (Kaminski, 2019). As such, explainability plays an important role in building relationships of trust to underpin a SL. Ensuring that the ways in which AI is used and decisions that are made based on data are understood, is crucial to facilitate the good communication and dialogue needed to establish a SL (Moffat et al., 2016).
Technical approaches to ensure explainability aim to demonstrate Ability through technical competence, while also enabling assessments of an organisation’s Benevolence and Integrity. Indeed, transparency is crucial to enable insights into an organisation’s motivations or values. Therefore, explainability may not directly demonstrate Benevolence or Integrity but might constitute an important feature to enable assessments of these characteristics. Moreover, explainability may be vital to facilitate public/stakeholder engagement and dialogue essential for establishing a SL.

**Lineage**

As AI models evolve and adapt, transparency can be problematic and the “black box” nature of AI is amplified. Approaches focusing on lineage aim to make the inner components and the history of the AI algorithm traceable by logging the necessary details and keeping track of the interactions occurring between components. The European Commission (2019) advocate traceability of AI algorithms to include documenting all the data sets and processes involved in the data gathering and data labelling phases. Traceability is regarded as essential to enable ‘identification of the reasons why an AI-decision was erroneous which, in turn, could help prevent future mistakes [while also] facilitate[ing] auditability as well as explainability’ (European Commission, 2019: 18).

Technical approaches enabling traceability of the lineage of AI models provide insights into how these processes have developed, placing an emphasis on transparency compared to explainability of current processes or outcomes. However, traceability also enhances explainability (European Commission, 2019) and perceived reliability/predictability. As with explainability, these approaches aim to demonstrate Ability through technical competence while also enabling assessments of an organisation’s Benevolence and Integrity through deeper forms of transparency.

**Trade-offs**

These four broad classifications illustrate the range of different technical approaches being developed and used to address ethical challenges relating to data practices and AI. While each of these may be important for developing trustworthy practices to underpin a SL, it may not always be possible to achieve each of the four aims simultaneously. Indeed, even within each of the approaches trade-offs may be necessary, for example fairness can have different meanings and be assessed via different measurements, therefore ensuring fairness may require prioritising certain methods which in turn prioritise different dimensions of fairness (Chouldechova, 2017). Additionally, the approaches outlined above are not always complementary to one another. There is significant interest in Explainability given the requirements brought in by GDPR; however, this poses challenges in relation to many AI applications (Goebel et al., 2018). Where a technology can be developed to be robust and fair but is not fully explainable, trade-offs may be necessary. Such trade-offs might have important implications for trustworthiness and for establishing or maintaining a SL. Therefore, understanding stakeholders’ interests and values in relation to the way their data is used or the ways in which technologies are deployed will be valuable to guide decision-making in these instances. Moreover, transparency around these trade-offs and the ways in which particular features of technologies have been prioritised may be important to maintain relationships of trust with stakeholders.

**Social approaches**

Developing technical approaches to address ethical challenges will be an important component to underpin a SL for current and future practices of FinTech organisations; however, technical solutions alone are insufficient to achieve this outcome (European Commission, 2019). As outlined above, a proliferation of frameworks and discourse surrounds the technical approaches through which to pursue trustworthy data practices, conversely, there is considerably less discussion of the social approaches needed to complement these – or how social approaches might be undertaken in data-intensive industries. Yet, social approaches are central in establishing a SL. The following section considers the implications of this for FinTech.

**Public engagement**

A SL is established and maintained through ongoing relationships between a community of stakeholders and an industry/organisation. This entails ongoing engagement and dialogue to identify and respond to stakeholders’ values, interests and concerns (Moffat et al., 2016).

Current levels of interest in public engagement with data practices are high (particularly relating to AI), reflected via the growing number of bodies working in this area (including Google DeepMind Ethics and Society, the UK Government’s Centre for Data Ethics and Innovation, the New Zealand Government’s Data Futures Partnership and the Royal Society). It is now widely recognised that Big Data analytics and AI bring significant social and economic impacts and necessitate both regulatory supervision and ethical and social assessment (Stahl and Wright, 2018).
Consideration of the social and ethical dimensions of data practices reflects a longer history of public engagement with science and technology. In the past, public engagement has been promoted as a means to address contentious areas of innovation and to build or restore public trust and mitigate controversy (Aitken et al., 2016a). This is deemed important as ‘science and technology demand assenting publics to maintain their hold on the collective imagination, not to mention purse-strings’ (Jasanoff, 2011: 248). However, public engagement goes beyond communicating the value of science and technology, and instead requires engaging in dialogue with the public to understand and reflect public values in innovation, governance and policy.

Previous studies in public engagement with science and technology have demonstrated the limitations of approaches aimed at gaining public trust through improving public understanding. Such approaches treat members of the public as ‘passive recipients of scientific knowledge’ (Cunningham-Burley, 2006: 206), overlooking how members of the public critically assess, deconstruct and evaluate claims to scientific knowledge in line with their own ideologies, experiences and the contexts in which the information is received (Hagendijk and Irwin, 2006). Thus, demonstrating technical competence or communicating the robustness of technical responses to ethical challenges will not automatically lead to public trust and support. Rather, technical approaches need to be combined with social responses that build relationships of trust through which claims to technical competence, and demonstrations of ABI will be evaluated. As such, aligned with the approaches taken to establish a SL, rather than aiming to manufacture public trust in science and technology, the focus of public engagement is to ensure that the trustworthiness of science and technology evolves through efforts to address and reflect public values (Aitken et al., 2016a; Wynne, 2006).

To date, deliberative public engagement relating to data has typically been undertaken by research organisations or public sector bodies (e.g. Data Futures Partnership, 2017; RSA, 2018). Important questions arise regarding whether private sector (non-research) organisations, such as FinTechs, can, or should, facilitate these processes. Community engagement is a key component of establishing and maintaining a SL in extractive industries (e.g. mining and forestry); however, this has been undertaken with varying degrees of commitment and quality (Moffat et al., 2016). In some cases, community engagement has been largely cosmetic due to companies retaining control over the process, restricting the range of possible outcomes and setting the terms for community participation: ‘even when all key stakeholders are explicitly invited into a conversation […] asymmetric power relations between parties, and differences in value sets, worldviews and perspectives are still likely to create opportunity for mistrust and conflict’ (Moffat et al., 2016: 483). These remain persistent challenges in public participation across a variety of domains and ones which are fundamental to address, in order for public engagement surrounding AI, and data-intensive industries such as FinTech, to be meaningful and impactful.

The potential motivations for FinTechs to undertake public engagement will shape the aims of engagement, the approaches taken and the range of potential outcomes. As noted above, there are a variety of reasons for FinTech firms to pursue a SL: these may range from purely instrumental perspectives which regard the SL and related public engagement as a mechanism through which to attract and retain customers and increase profits, through to substantive perspectives which focus on bringing wider benefits for society and meaningfully involving members of the public to address ethical considerations. Clearly such rationales lead to different approaches being taken and different ideas of what it would mean for public engagement to be successful (Aitken et al., 2016b). Experience in other industries suggests that approaches informed by instrumental rationales may have the most appeal to private sector organisations, yet those informed by substantive rationales are more likely to be effective (Aitken et al., 2016b). For example, a review of community engagement practices by wind farm developers found that while most developers took an instrumental approach to community engagement (using methods which restricted the ways people could participate or the range of potential outcomes), those that followed more substantive approaches (opening up engagement processes and devolving some control over the process and outcomes to public participants) were most successful in generating public support, which was ironically the primary objective of companies following instrumental approaches (Aitken et al., 2016b). Therefore, while companies may be reluctant to share power in decision-making or planning processes, evidence suggests that doing so leads to positive outcomes in terms of generating wider public support and establishing a SL.

As noted above, a SL is granted through engagement with ‘local community stakeholders who are affected by [a project or development] and those stakeholders who can affect its profitability’ (Moffat et al., 2016: 480). In extractive industries, identifying ‘local community stakeholders’ may be more straightforward given the physical location of projects. For FinTechs, a ‘local community’ defined by geographic proximity is in most cases irrelevant to the operations of a FinTech. Instead, while in extractive industries local communities have been identified based on physical proximity to
the locations from which resources are extracted, in
data-intensive industries affected communities might
be conceptualised as those from whom data is derived.
This creates a much wider set of relevant stakeholders.
Moreover, taking a broader approach to stakeholders
as people who are “affected by a project” necessitates
consideration of the impacts of data practices on
society. Such impacts might include potentially
transformative effects on financial systems which affect
people’s access to finance (either positively or negatively)
or contributory effects to the increasing role of data in
society and the reduction in opportunities to partici-
pate fully in society without allowing one’s data to be
collected or used. Such a broad conceptualisation sug-
gests that stakeholders might include the whole
of society.

Considering the second group of stakeholders as
those ‘who can affect profitability’ may also invite
either narrow or broad definitions. Narrow definitions
might focus on potential and actual customers as those
who can affect profitability. Broad definitions would
consider the role of the wider public as potentially
affecting profitability through their support or opposi-
tion to data practices more broadly as well as those
used specifically in FinTech. Indeed, as previous scan-
dals have demonstrated, public controversies around
data use and misuse have the potential to significantly
affect data dependant industries (as was evidenced in
the case of Care.Data (Carter et al., 2015)).

Therefore, establishing and maintaining a SL entails
going beyond a narrow focus on stakeholders as a com-
pany’s customer base to a more inclusive conceptualis-
ation of the wider public as stakeholders. Yet, the
extent to which a private sector organisation – such
as a FinTech – will be willing, or adequately resourced
to engage with such broad stakeholders is questionable.
Instrumental approaches to engagement will be likely
to lead to a narrow focus on stakeholders as existing
or potential customers (those considered to have the most
immediate impacts on profitability); however, over-
looking wider stakeholders risks practices leading to
unanticipated negative impacts or opposition to
approaches which are not aligned with public values.
Therefore, a FinTech may not be granted a SL for its
operations if it overlooks the interests of broader stake-
holders. This highlights that while a FinTech may
define its stakeholders in particular ways, others
(including stakeholders themselves) might define them
differently and it is the stakeholders rather than the
FinTech which has the authority to grant, refuse or
withdraw a SL for its operations. Thus, taking a
narrow approach to defining stakeholders may be a
short-sighted and risky strategy.

Nevertheless, taking a broad approach presents fur-
ther challenges. While organisations such as the
New Zealand Government, the Royal Society or
DeepMind Ethics and Society have substantial budgets
and resources which they can use to fund large scale
public engagement projects to reach out to diverse
groups across society, FinTech companies are unlikely
to have significant resources (or expertise) for these
activities. Furthermore, given that the SL for FinTech
is interdependent with a SL for broader data-intensive
industries, and innovation, questions arise as to who is
responsible for facilitating engagement activities.
Individual FinTech firms have an incentive to develop
a SL for their own operations, yet, it may be that wider
industry level engagement is needed to establish a
broader SL for the FinTech sector.

Indeed, public engagement can occur at a range of
levels reflecting different aims and objectives and
requiring different approaches. For example, public
engagement relating to data-intensive health research
takes place at many different scales, including: ‘wides-
scale public conversations about uses or potential uses
of data in health research; [public engagement] to
inform or co-design the development of policies or gov-
ernance practices relating to uses of data in health
research; engagement or involvement of members of
the public in governance decisions about data access
and use; engagement or involvement of members of
the public at different phases in particular research
projects; analysing and disseminating the results of
research using data in ways which will support
improvements in healthcare and systems’ (Aitken
et al., 2019: 2). Moving this approach into the
FinTech context suggests public engagement might
valuably serve a similar range of purposes: at times
being undertaken at policy or industry level to inform
the development of policies, governance mechanisms
and industry practices, and at other times being under-
taken by individual FinTech firms to address ethical
dimensions in developing, implementing and evaluat-
ing new products, services or areas of innovation.

Conclusions

Despite the substantial and growing rhetoric around
ethical and trustworthy data practices (in all sectors)
there is limited evidence of how this is being put into
practice. As this article has discussed, this is important
for FinTechs who increasingly employ these technolo-
gies to underpin new financial services. While there is
industrial advocacy surrounding the potential benefits
of data science and AI in banking, it is not yet clear
whether there is a SL for these practices.

This has wider implications for developing ethical
data practices. The proliferation of ethical codes of
practice and guidance well-illustrates that ethical prac-
tice requires more than just regulation. However, it is
debate whether the growing number of codes of practice is in reality leading to meaningful changes. In particular, given that such guidance are beyond regulation they are enacted voluntarily with little or no enforcement. This means they depend on organisational culture change to realise their value. Such culture change in turn requires meaningful commitments to ethical practice from senior levels of management. There may be a range of motivations for organisations to adopt ethical codes of practice; however, we argue that framing this in terms of pursuing a SL for the operations of data-intensive industries provides a clear rationale and set of approaches to underpin emerging ethical best practices.

SL is distinct from approaches such as CDR (Cooper et al., 2019; Lobschat et al., 2020) in that it places public – or stakeholder – engagement at its heart. Since a SL is granted or refused by external stakeholders (rather than secured internally) it focuses attention at the importance of aligning with public values through public engagement. Enforcement does not come through formal sanctions or penalties but rather through the loss of public trust, legitimacy or credibility which can have substantial and far-reaching implications for an organisation and industry.

Establishing a SL underpinned by relationships of trust requires FinTechs to combine a range of technical and social approaches and continually reflect on ethical dilemmas as well as the extent to which practices align with social values. In this regard, the growing body of guidance and best practice regarding responsible or trustworthy AI, ethical data practices and CDR represent a valuable set of resources to draw upon, yet it is important that this goes beyond rhetorical commitments and leads to practical and meaningful action. In particular, establishing trustworthiness requires not just demonstrating technical competence (or Ability) but also Benevolence and Integrity in the ways that data is used and technologies are deployed. Moreover, in order to align with public values, it is vital that ethical approaches are informed by the views and interests of broad stakeholders. In the case of FinTech, establishing a SL for these technologies and subsequent services may prove vital to the ongoing success and sustainability of this sector.

FinTech firms face a number of challenges in establishing relationships of trust: First, the damaged reputation of the financial sector as a whole (Dietz and Gillespie, 2012). Second, the unfamiliarity of technologies driving FinTech products and services. Third, increasing public awareness of controversies around data misuse and, fourth, that as new entrants to the financial marketplace FinTechs have yet to establish widely recognised brand reputations. On the one hand, this ‘newness’ may offer FinTechs a competitive advantage as an alternative to traditional banking incumbents (King, 2018). On the other hand, it means that there may be substantial work required to establish relationships of trust with the wider public. Yet, there is also an opportunity to develop new approaches which might further enhance competitive advantage. As has been noted by Brusoni and Vaccaro (2017: 223) ‘the ethical standing of an organisation—that is represented by its internal practices, products and services—clearly provides a unique way to differentiate from competitors’.

This article has not aimed to identify public interests or concerns relating to data practices in FinTech, or to set out what is required for FinTech to align with public values. Since there is a paucity of public engagement or deliberation examining public values around FinTech practices, further research (including through public engagement methods) is needed to examine what this means in practice. Therefore, this article focuses on setting out the approaches needed to achieve this.

We posit that such approaches are needed across all areas and industries whose operations are dependent on data to complement regulation and inform the development of ethical codes of practice. This is important to underpin culture change and to move beyond rhetorical commitments to develop best practice putting ethics at the heart of innovation.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the EPSRC, grant reference: EP/R033595/1.

ORCID iDs
Mhairi Aitken https://orcid.org/0000-0002-4654-9803
Karen Elliott https://orcid.org/0000-0002-2455-0475

References


Schueffel P (2016) Taming the beast: A scientific definition of Fintech. Available at SSRN 3097312.


The Economist (2017) The world’s most valuable resource is no longer oil, but data. Available at: https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data (accessed 9 May 2019).


