An Updated Model of Trust and Trustworthiness for the use of Digital Technologies and Artificial Intelligence in City Making

Brydon T Wang
School of Law, Queensland University of Technology
Brydon.Wang@qut.edu.au

ABSTRACT

Although digital technologies that inform city making are seen as tools to enhance democratic processes in the decision-making inherent in placemaking, these technologies are critically dependent on trusted relationships and play a role as trustworthy channels of communication between urban planners, policy makers and communities. This paper examines how the existing use of digital technologies in city making signals trustworthiness in both the design and deployment of such technologies, and considers how the increasing use of artificial intelligence ('AI') will require careful implementation of benevolent transparency processes. The paper presents an updated model of trustworthiness for the use of digital technologies and AI in city making based on Mayer et al's (1995) widely-cited integrative model of trust (the 'ABI model'). It begins with an analysis of the need for greater demonstrations of trustworthiness as digital technologies are equipped with AI. The paper then introduces the ABI model, breaking down the components of trust in interpersonal contexts to focus on the perceived factors of trustworthiness: *ability, benevolence* and *integrity*. Ability considers the technical competence of the recipient of trust (the 'trustee'). In the context of city making, this trustee would be the urban planner, policy maker or technology developer, and their ability to design and deploy these digital technologies and AI in city making. Benevolence considers whether or not the trustee holds a positive orientation or intention towards the member of the community who is the intended giver of trust (the 'trustor'). Integrity considers whether or not the trustor and trustee subscribe to a common set of values. This paper argues that the germinal component of trustworthiness lies in signals of benevolence as articulated in transparency practices that demonstrate mutual vulnerability, consensus formation and seamful decision-making processes. These benevolent qualities of transparency practices are mapped onto existing uses of digital technologies in city making to demonstrate how the updated model of trust and trustworthiness can help visualise the signals of trustworthiness as AI is increasingly embedded in these digital technologies.

MAB '23, June 14–23, 2023, Toronto, ON, Canada

ACM ISBN 979-8-4007-1635-5/23/06
https://doi.org/10.1145/3627611.3627618

CCS CONCEPTS

• **Human-centred computing**; • **Applied computing**; • **Theory of computation**;

KEYWORDS

Trustworthiness, Data Governance, Urban Interaction Design, Digital Placemaking

ACM Reference Format:
Brydon T Wang. 2023. An Updated Model of Trust and Trustworthiness for the use of Digital Technologies and Artificial Intelligence in City Making. In *Proceedings of the Media Architecture Biennale 2023 (MAB '23), June 14–23, 2023, Toronto, ON, Canada.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3627611.3627618

1 INTRODUCTION

Increasing datafication of the city through the use of digital technologies for city making has led to broader collections of data at increasing volumes and granularity. As the collected data is analysed to achieve greater visibility and understanding of the data subject [1], the accompanying appetite for these data-focused digital technologies and 'AI to design and implement city management strategies' [2] necessitate a deeper consideration of how such systems are, at once, dependent on trustworthy practices and simultaneously support signals of trustworthiness in the community.

This paper focuses on Byrnskov et al's definition of media architecture as the 'design of physical spaces at architectural scale incorporating materials with dynamic properties that allow for... reactive or interactive behavior' [3]. It takes as media architecture the digital technologies that transform buildings into Architectural User Interfaces ('AUI's) [4], and includes data analytical tools and algorithms that sit behind these AUIs that are currently in tension with the technology-driven participatory culture of media architecture that has developed over the past decade-and-a-half [5]. Given this tension, this paper argues that trustworthy practices need to be designed into both the design process and eventual deployment of algorithms and other data analytical tools in media architecture and urban interaction design. This is because the design of workshop processes that support public acceptance of media architecture is heavily dependent on trusted relationships [6]. Where digital technologies are designed through trustworthy processes and deployed to serve as a 'credible and trustworthy... channel for civic dialogue' [7], such media architecture technologies reduce 'barriers between government and communities', allow for experiential understanding (through augmented reality or full-body interactions) of proposed changes, and enhance transparency in city-making processes [7].

Transparency, in turn, allows for parties involved in decision-making to be made accountable through visualisation of the data

lifecycle: how data is collected, stored and used to drive decision-making in the urban context. Similarly, when these media architecture technologies are deployed, transparency is required to '[show] information in an objective, fair, and trustworthy way that accurately reflects the actual situation. For instance, the sources of the data that were used are clearly mentioned and are publicly accessible' [8]. In this manner, transparency as a key practice of city making allows stakeholders to examine and critique the 'technological agenda of "seamlessness"'[9], particularly where AI is used with digital technology to move seamlessly from data collection point to decision output. Accordingly, transparency is crucial to how trustworthy, and subsequently, how effective the relevant media architecture technology will be as a digital placemaking tool and means to achieve human-centred smart cities [10].

This paper examines how media architecture technologies can demonstrate trustworthiness in the design and deployment of such technological products. It seeks to advance theoretical considerations of trustworthiness in data governance to contribute to the regulation and accountability of media architecture and the use of technologies in the development of cities. This paper begins with a brief exploration of data collection practices in the increasingly datafied city that necessitates greater clarity in how we think about trustworthy data practices. It then briefly outlines the definitional challenges presented by the terms 'trust' and 'trustworthiness' that impede clear decision-making around how we collect and use data to drive decision-making in the city. In examining these conceptions of trust, the paper then introduces the integrative model of trust and trustworthiness ('ABI model') that has been extensively deployed across various disciplines and fields of research, examining how the ABI model has been applied to digital technologies and AI tools. Significantly, the paper then presents an updated model of trust and trustworthiness and articulates the germinal nature of benevolence as a signal of trustworthiness in demonstrations of transparency. It then applies the updated model of trust and trustworthiness to various contexts where digital technologies are used in city making to highlight how benevolent transparency practices are extant in media architecture but need to safeguard against the increasingly opaque nature of AI applications. Finally, a conclusion is presented as to the utility of the updated model in better understanding how digital technology and AI can be used in city making in benevolent, trustworthy ways.

2 TRUSTWORTHINESS IN THE CONTEXT OF THE DATAFIED CITY

Data collection is crucial to use of technologies for decision-making in the increasingly datafied city. Such decision- making requires the aggregation of relevant and diverse data selected against the criteria for the decision output [11]. However, the timeliness of data collection affects the types of decisions that can be made based on such data. This need for real-time data collection is being increasingly met by the 'growing ubiquity' [11] of sensors embedded in our built environment, through media architecture [12], [13] or carried on mobile devices. These various sensing technologies extract data from our digital trails and sit as nodes within larger sensor networks [14], all connected by the internet of things ('IoT'). Data collected through these sensors in the built environment produces a constant stream of 'urban data' [15], [16] which then feeds into the data architecture of various automated decision-making systems. Miniaturised sensors are now embedded in the fabric of our cities through smart lampposts [17], [18], or carried dynamically through our streetscapes and urban room in ubiquitous objects like mobile phones, pens, and wearable devices [19]. The resultant sensor-generated data offers a glut of information about our everyday lives (both in the physical and digital worlds). We can interrogate this information in any number of ways and repeatedly at will to decipher fine-grained, street-level actions and predict patterns of behaviour that are scalable from the individual building up to an entire city [20].

The tracking of people within the urban environment offers other benefits and challenges. There is a long practice of using collected data from the city, such as data derived from 'censuses, household, transport, environment and mapping surveys, . . . commissioned interviews and focus groups' to support and track the efficacy of urban interventions [21]. In the 1960s, William Whyte's 'Street Life Project' articulated a germinal project of mechanically collected urban data through video cameras installed within the streetscapes of New York City to film and record pedestrian movement [22]. This systematic and granular record of interactions of pedestrians informed New York's 1969 city plan, underscoring the key role mechanically-originated data plays in rendering urban residents and urban conditions 'visible' and enabling governments to 'see' [23], [24]. Scott conceptualises this type of 'seeing' as statecraft directed at making populations 'legible'. In this manner, individuals within society are rendered 'legible' through the collection and consumption of data that the state obtains on 'its subjects, their wealth, their landholdings and yields, their location, their very identity. . . [to create] a detailed "map" of its terrain and its people [25].

While there has always been an intimate link between data-generation practices and city development, new forms of sensorised data generation are changing notions of city life that are oriented towards 'smarter' outcomes [26]. The emerging global imperative to create 'smart cities' has seen a push to embed or retrofit sensors in new buildings and existing brownfield sites [27]. Sensorised technologies in the built environment have also captured the imagination of proponents of participatory urbanism [28]. With the promise of new and larger datasets to help inform decision-making around the city [29], proponents seek to achieve 'superior levels of performance, new forms of functionality, transparency in allocation of resources, and good economics over long time horizons' [30]. In doing so, they seek to harness the 'participatory uses of technologies, big data, and social media. . .[to] inspire new smart city design approaches for urban designers' [28].

This has seen an acceleration of new data analytical models employed in city making. Data collection now occurs 'through sensors, connected objects (IoT), interfaces, travelling cards, smart meters, mobile apps, implants, cameras, e- government systems, polls, online platforms and more' [23]. Data from sensor technologies, such as UAVs conducting low altitude aerial photogrammetry [31], have been combined with traditional land-surveying techniques to automate field mapping and surveying of cities [32], [33]. Terrestrial laser-scanning technology (a high-precision 3D measurement technology) has been combined with Global Positioning System

('GPS') data to produce synchronous point cloud data that could accurately process urban morphology [34]. Utility networks are being modelled using BIM (Building Information Model) data, geographic information systems ('GIS') and graph-theory abstractions [35]. In media architecture, we are seeing data as video and audio recordings that are used to 'examine the emplaced experiences of. . . participant groups' to bring new knowledge of how we 'experience dynamic adaptation, informational services and interactivity in the built environment' [36]. In other scenarios, sensor-derived data from the built environment have been combined with data drawn and crowdsourced from 'social media platforms such as Twitter, or location-based services' [37], to provide us with a deeper, granular understanding of the affective experiences of urban residents [38].

However, these new forms of seeing through urban data are accompanied by challenges. While these forms of data add to the breadth of heterogeneous content arising from the various disciplines and ways of measuring within different contexts, such diversity in data collection practices results in fragmented information involving 'behaviors and control that are distributed and concurrent' and interdependent within the urban environment of the city [30]. More importantly, the types of sensors that are being developed and deployed both in buildings and the wider context of the city are intimately linked to the problems a decision-maker intends to solve. This exercise in defining the problem—where an abstract goal is translated into a 'predictive goal' and then into a 'specified outcome variable' [39]—is crucial as a scoping exercise to designing the types of data sought, data collection approaches, and how the data is intended to add value to the decision-making process. The impartation of value into 'data extracted and abstracted from reality' describes the 'datafication' that defines smart cities [23], but is accompanied by the coding in of opinions and bias into digital processes that are then used to make decisions that make a city. By attributing value to data, data becomes essential to the functions of a smart city and in this way, 'the smart city bases its reality on data, and turns that data back into its reality' [23].

Given the permeation of data practices in decision-making around the city, the construction of what data is and what data is then collected and sent through the decision-making systems requires examination. Bias can thwart the effectiveness of data collection [40], [41]. Critically, where bias is built into the data and the various digital layers data passes through, deep systemic issues arise that erode trust relations in the city. Hence, it is crucial to ensure that bias does not enter the scoping or design process, particularly when 'defining the output variable and labelling its constituent classes, collecting and labelling the training data, and selecting the input variables' [39]. Likewise, the insertion of media architecture as a mechanism by which 'processes of commercialization and mediation' [42] intrude and colonise public spaces need to be considered, particularly as we use data obtained from these urban interventions to drive decision-making around cities.

This has led to questions of whether data practices in the city—how we use data to drive decisions about how and where we locate and provide services and infrastructure in the city ('digital placemaking') and involve public participation in such decision-making processes ('urban interactive design')—can be made trustworthy. These questions of what is trust and trustworthiness are situated in a larger context of the significant erosion of trust in institutions that has occurred over the past four decades that began with Robert Putnam's widely cited *Bowling Alone: America's Declining Social Capital* [43]—that examined the impact of the decline in those connections between individuals and groups on democratic systems [44]; the financial collapse and scandals of Enron and the dot-com bubble [64]; and most recently, the *Cambridge Analytica* scandal [45], [46]. This context and the acceleration in the use of data-driven automated decision-making systems in our everyday life in cities [47], [48], thus underscore the need to think and work through the trustworthiness aspects of the design and deployment of media architecture.

3 CONCEPTUALISING TRUST AND TRUSTWORTHINESS

The concept of trust is difficult to define or measure [49–52], Trust transcends the 'territorial barriers of social disciplines' [53], which has led to a plethora of specialised research definitions and contributed to 'literature confusion' [54]. Roger Mayer, James Davies and F. David Schoorman argued in their seminal work that research on trust is 'hindered . . . [by] a lack of clear differentiation among factors that contribute to trust, trust itself, and the outcomes of trust' [55]. This lack of clarity has led to a tendency to conflate trust with 'faith' and 'belief' and to a difficulty in establishing a measurement for trust—that is, whether to measure it against 'distrust', 'mistrust', or merely diminishing degrees of trust that eventuate in a total lack of trust [56].

The resulting 'conceptual morass' [57], [58] means that trust is defined in fundamentally different ways. It can be described as a personality trait or construct [59], a rational choice [60], a characteristic of interpersonal relationships [61], or a phenomenon linked to social contexts [62]. Simpson goes as far as to assert that '[t]here is a strong prima facie case for supposing that there is no single phenomenon that "trust" refers to, nor that our folk concept has determinate rules of use' [51]. Unsurprisingly, the definitional challenges mean that research into trust has to examine the different everyday contexts that shape varying conceptions of trust across distinct academic disciplines.

The differences in these varied definitions of trust are not 'trivial' [63]. Conceptualising trust requires balancing different disciplinary perspectives while ensuring that the framework does not become 'inordinately abstract' [63]. The framework needs to be grounded in everyday practice to avoid the concept of trust becoming too vague [54]. However, it is common in the literature to find that the framing of trust is used as a backdrop to calibrate reader expectations on conceptual overlaps and the complexities in the 'multiple forms' trust takes in 'the varied conditions of its exercise' [56].

To avoid the 'conceptual morass', this paper applies the model of trust proposed by Mayer et al: the ABI model (also known as the 'Integrative Model of Organizational Trust') [55] and focuses on the component factors of perceived trustworthiness. The model, which distils insights from the trust literature drawn across several disciplines—namely, 'management, psychology, philosophy and economics' [64], was initially designed to describe the formation and re-formation of trust in organisational settings [55]. However, because the model is drawn from multiple social disciplines, it has

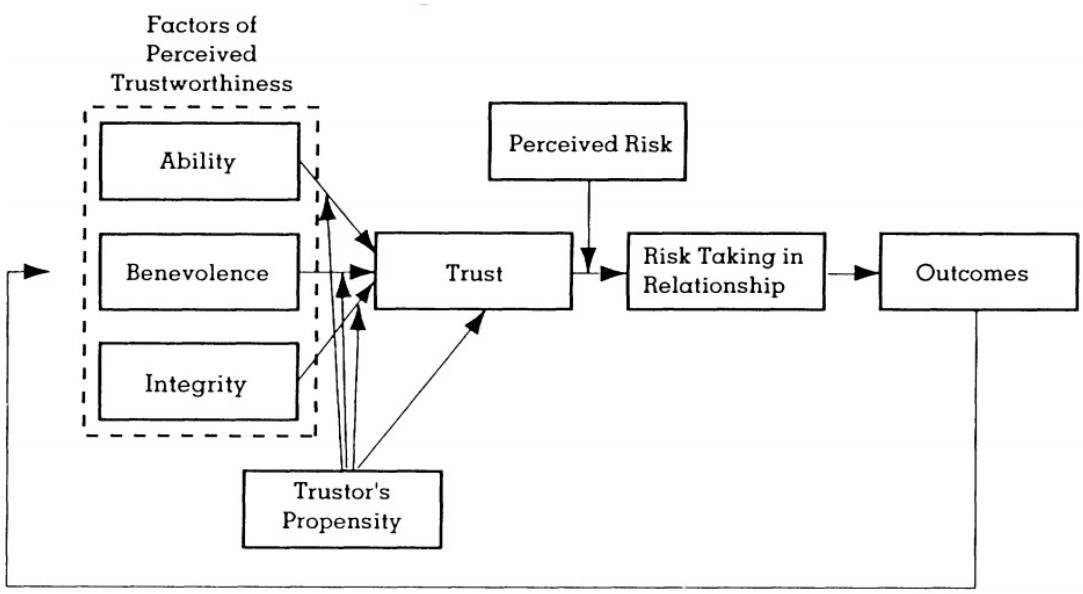

Figure 1: The ABI Model of Trust [55]

broad appeal and has been widely used across numerous domains of expertise [65–68].

The ABI model presents interacting components that form trust and trustworthy behaviours within one-to-one ('dyadic') relationships between trustee (the intended recipient of trust) and trustor (the person giving trust). The model focuses on the component elements that comprise the decision-making process a trustor undergoes when deciding to trust. These interlinked components establish a process for establishing trust, as outlined in Figure 1.

The model's process of trust formation has five components:

- *Trust as a relational construct.* The model involves trust formation in relationships—namely, the propensity or disposition of a person giving trust ('trustor') and the perceived trustworthiness of a potential recipient of trust ('trustee').
- *A trustor's propensity.* Trust as a *trusting attitude* only emerges as a function of the trustor's propensity to trust and the perceived trustworthiness of the trustee.
- *The factors of perceived trustworthiness.* The propensity to trust is based on three factors of perceived trustworthiness—ability, benevolence and integrity.
- *Perception of risk and vulnerability.* The propensity to trust derives from the perceived risks arising from the context of the trust scenario.
- *Risk-taking in a relationship.* Risk-taking within a trusted relationship gives rise to *'trusting behaviour'* and emerges from the decision-making process that balances perceived risks against the trustor's willingness to be vulnerable to risk. The outcome of the trusting behaviour—whether or not the trustee behaves in line with the trust expectations of the trustor—depends on a feedback loop that moderates future perception of a trustee's trustworthiness.

The five components of the model interact as variables in one instance of a trustor's decision-making process. The components involve the initial formation of an attitude of trust towards a trustee, followed by the trustor's decision to engage in trusting behaviours. The willingness of both parties to trust underpins the application of all components, and it is no surprise that the model's widely-cited definition of trust is founded on this notion, namely:

> the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party [55].

Trust as 'willingness' articulates trust as a trusting attitude distinguishable from the outcomes of such an attitude: the decision to engage in 'trusting behaviour' demonstrated by a resultant act of reliance [69]. In the ABI model, the decision to engage in trusting behaviour (or risk-taking) is a function of trust and the perceived risk in a given context. Therefore, in this single instance of the process, trust forms as a trusting attitude midway through the model before an assessment of the risk. Then a decision on whether or not to engage in risk-taking (as 'trusting behaviour') is taken.

The ABI model enunciates three factors of perceived factors of trustworthiness—*ability*, *benevolence* and *integrity*. A trustee communicates ability through clear demonstrations of competence, benevolence in showing a clear positive orientation towards the trustor, and integrity where the trustee demonstrates their commitment to adhering to a value set shared between trustor and trustee. While this model explores the antecedents to trust formation, this paper focuses primarily on the qualities of a potential recipient of trust (the trustee). Further, where this attitude of trust is applied to the design and deployment of digital technology in the urban context, this paper argues that these remain social relationships

mediated through a technological interface. Thus, the factors of perceived trustworthiness within the ABI model are used in this way to examine how trustees demonstrate ability, integrity and benevolence to increase the willingness of a potential trustor to be vulnerable to the risk of digital technologies deployed in city making in ways that could be detrimental to them.

4 APPLICATION OF THE ABI MODEL TO DIGITAL TECHNOLOGIES AND AI IN CITY MAKING

The ABI model has been utilised across multiple domains of expertise. Its development paralleled other notable works on trust [70–72] that emerged in the public debates in the mid-1990s on the erosion of social capital and trust in American institutions [73]. The model has also been used to scaffold other conceptual frameworks that seek to consider trust formation in the context of human-to-machine trust [74].

Palmer et al examined the role of intermediaries in developing trust on the internet in the late 1990s [75]. They applied the ABI model to the relationship between human consumers and 'Internet firms' through the intermediary of a computer interface. The researchers observed that these 'Internet firms' needed to communicate online, and they established trustworthiness via the intermediation of the internet's infrastructure [75]. Their observations remain highly relevant in the context of media architecture and the broader context of digital technologies deployed in city making.

Palmer et al considered that in the context of e-commerce, these firms had to demonstrate the three trustworthiness factors by communicating:

- *ability* to the consumer or data subject—ie that the firm or technology developer had specific 'skills and competencies to deliver services' [75];
- *integrity* to the consumer or data subject by showing they had governance policies that are deemed by the consumer or data subject to be appropriate. One suggestion was a privacy policy that sets out how data would be collected and handled by the firm or technology developer [75]. In the case for media architecture practitioners, clear policies and the adoption of voluntary schemes that govern the use of AI in city making would be critical to establishing value congruence with users and providing vertical accountability;
- *benevolence* to the consumer or data subject by showing that their business practices went beyond a pure profit- seeking model to balance the data subject's interests, welfare and needs [75]. However, discussions on benevolence (including Palmer et al's translation of benevolence) are ill-defined and do not offer clear directions that will allow an individual data subject to feel a personal orientation that means them well.

Palmer et al's translation of interpersonal or inter-organisational dyadic trust relationships to the trust context between humans and machines through the intermediary of technology echoes a similar approach taken by Bernard Barber. Barber suggested three elements of human 'expectations' to serve as the basis for how a human decides a machine is trustworthy: technical competency (ability), persistence (integrity), and fiduciary responsibility (benevolence) [57].

With the entry of AI and other algorithmic systems in media architecture, there is a need to test the continued applicability of the ABI model to the trust scenarios inherent in each interaction with media architecture technology. Thiebes et al offers one of the more compelling translations of the ABI model for digital technologies [76]. Table 1 below traces how the ABI factors of perceived trustworthiness in scenarios of interpersonal trust have been translated into new models for trust scenarios involving digital technology, and then automated and autonomous systems using AI. Their proposal for five trustworthy AI ('TAI') principles based on their survey of the iterations of the ABI model provides an example of an emerging value consensus that underpin expressions of benevolence in the design and deployment of AI, and the wider context of automation technology and autonomous systems. This consensus, in turn, then drives the production of set standards for integrous behaviour and demonstrations of ability.

However, this paper argues that the original factors proposed by Mayer et al remain robust and more suited to an analysis of trustworthiness in digital technologies and AI deployed for city making than these newer translations. For example, in relation to the terminology proposed by Thiebes et al of 'performance', 'purpose', 'process'— these terms lack the clarity of the ABI model and the factors of perceived trustworthiness.

- *Benevolence* more clearly evokes the social purpose of the design and development of Trustworthy AI ('TAI'). However, unlike the more generic term 'purpose', benevolence targets the benefits of the technological system at the individual level.
- *Integrity* then focuses on the process of technological development and its adherence to the purpose and intent of the TAI development. While Thiebes et al describe 'process' as a 'trusting belief' in the automation technology's algorithms being 'appropriate for the situation and able to achieve the [powerholder's] goals', this does not capture the value congruence that needs to be evident in the developmental process to demonstrate integrity.
- Finally, *ability* is demonstrated as the technical performance of the automation technology and autonomous system. To a certain extent, Theibes et al's description of 'performance' satisfies the translation of the factor of ability in scenarios involving 'trust in technology'.

Together, the original ABI factors of perceived trustworthiness are more suited to capturing the developmental requirements needed for digital technologies, particularly where they incorporate automated decision-making functions, to demonstrate trustworthiness. However, the terms 'purpose', 'process' and 'performance' are helpful in that they convey developmental stages. Technology developers and city authorities need to establish the 'purpose' for the media architecture technology before working through the developmental 'process' to achieve their 'performance' goals. Thiebes et al's translation of the ABI factors of trustworthiness to these terms highlights the germinal quality of benevolence. The significance in demonstrating benevolence in AI, media architecture and more generally trusted technologies of automation, is underscored

Table 1: 'Overview of common trusting beliefs' related to persons and technologies extracted from Thiebes et al [76]

Trust in persons (e.g., Mayer et al. 1995; McKnight et al. 2002).	Trust in technology	
	Trust in IT artifacts based on system characteristics (e.g., McKnight et al. 2011; Thatcher et al. 2010)	*Trust in automation technology and autonomous systems* (e.g., J. D. Lee and See 2004)
Competence / Ability: One has the ability to do for the other person what the other person needs to have done (McKnight et al. 2002). Group of skills, competencies, and characteristics that enable a party to have influence within some specific domain (Mayer et al. 1995).	**Functionality:** The belief that the specific technology has the capability, functionality, or features to do for one what one needs to be done.	**Performance:** The competency or expertise as demonstrated by the automation's ability to achieve the operator's goals.
Benevolence: One cares about the welfare of the other person and is therefore motivated to act in the other person's interest, does not act opportunistically toward the other.	**Helpfulness:** The belief that the specific technology provides adequate and responsive help for users.	**Purpose:** The degree to which the automation is being used within the realm of the designer's intent.
Integrity: The extent to which a trustee adheres to a set of principles that the trustor finds acceptable.	**Reliability / Predictability:** The belief that the specific technology will consistently operate properly (McKnight et al. 2011) and its behavior can be forecast (Thatcher et al. 2010).	**Process:** The degree to which the automation's algorithms are appropriate for the situation and able to achieve the operator's goals.

in the emergence of various processes to achieve social consensus on how these technologies can demonstrate a personal orientation 'motivated by care and concern to protect the trustor's [in this case, the data subject's] interests' [77]

Thiebes et al's difficulty in translating the trustworthiness factor of benevolence in trust scenarios involving automation and autonomous systems underscores the challenge in conceptualising design and developmental processes that demonstrate positive personal orientation towards the individual human in human-machine trust scenarios. However, in the context of participatory placemaking and the use of digital technologies for city making, we have clear examples that demonstrate this positive personal orientation, that are built to signal trustworthiness at the level of benevolence.

5 AN UPDATED MODEL: BENEVOLENCE, INTEGRITY AND ABILITY

This section of the paper sets out the author's updated model of trust and trustworthiness for digital technologies and AI in media architecture and city making. Figure 2 presents the updated integrative model.

The updated model reorders the factors of perceived trustworthiness from *ability, benevolence, integrity* to *benevolence, integrity, ability* because *benevolence* is the germinal factor of trustworthiness. It is benevolence that produces the set of values by which a trustee can demonstrate integrity. Likewise, the standards, regulatory frameworks and social norms that define integrous behaviour create the technical specifications and briefs that drive the performance of ability.

Starting from the left of the updated model within the trust formation stage, the trustor's propensity to trust comprises two components: 'Faith in Humanity' and 'Trusting Stance'. Where a Trusting Stance is adopted and overrides any consideration of the trustee's trustworthiness or perceived risk level, a 'default decision to trust' emerges as therapeutic trust. The updated model does not accept any risk-taking resulting from the therapeutic trust as an indicator of trusting behaviour. Where the Trusting Stance does not override considerations of trustworthiness and risk, both Trusting Stance and Faith in Humanity affect the trustor's ability to 'perceive' the trustworthiness factors. As mentioned, these trustworthiness factors have been reordered to reflect the germinal nature of benevolence.

In trust formation, a trustor's propensity to trust affects their ability to perceive the factors of trustworthiness to produce trust as a trusting attitude. The interaction of these factors and the trustor's inherent tendency to trust produces trust as a quantifiable willingness to be vulnerable to *'x' level of risk*.

Moving to the right of the model, in the subsequent segment on the decision to engage in trusting behaviour, the trustor's trusting attitude interacts with the perceived risk factors (*'y' level of risk*). Where a trustor's risk appetite exceeds the perceived level of risk ($x > y$) in the trust scenario, the trustor may decide to engage in risk-taking in the relationship. This will emerge as trusting behaviour. Where the perceived risk exceeds the risk appetite for risk ($y > x$), the trustor may choose to gamble and take risks in the relationship (but this is not risk-taking arising from a decision to engage in trusting behaviour), or they may choose not to accept any risk.

The outcome of any risk-taking (or lack thereof) is then used to moderate the factors of perceived trustworthiness. In exceptional circumstances, the feedback loop may also affect the trustor's propensity to trust; however, to do so, it must overcome the historical, cultural and sociological factors that influence the trustor's Faith in Humanity and Trusting Stance.

6 BENEVOLENT TRANSPARENT PROCESSES IN DIGITAL TECHNOLOGIES AND AI IN CITY MAKING

This paper now turns to the factor of benevolence and its essential role in allowing the use of digital technologies in city making to demonstrate trustworthiness. It demonstrates that existing transparent practices of participatory placemaking operate to send

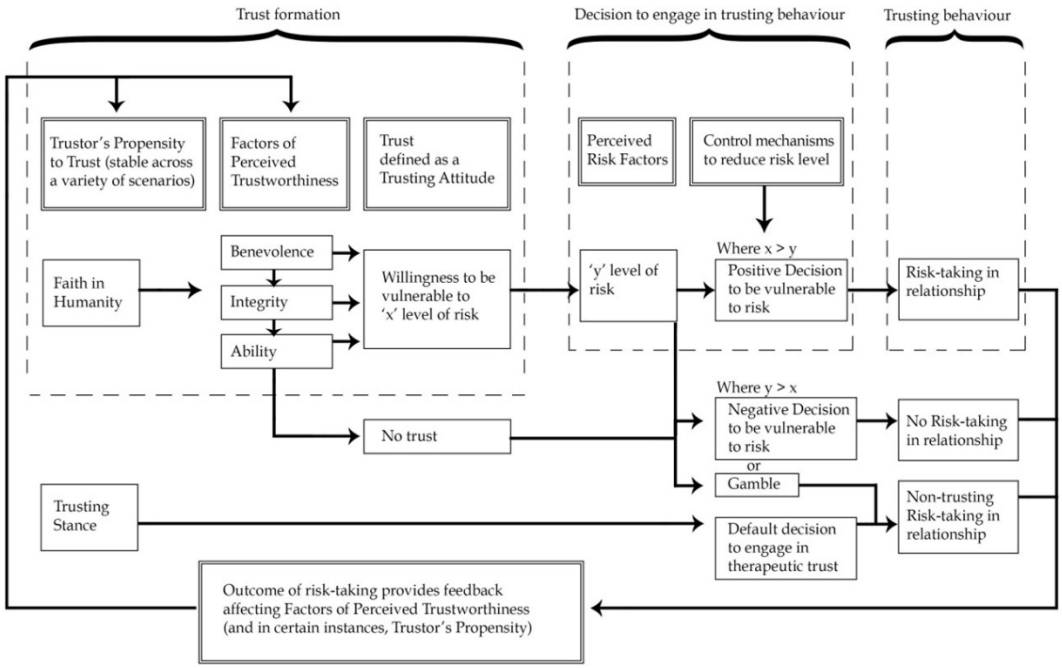

Figure 2: Updated Integrative Model of Trust and Trustworthiness

signals of trustworthiness from the urban designer or technology developer to a participant assessing if they can trust the design process and the deployment of such technologies. As observed in the preceding sections, the trustworthiness factor of benevolence takes on a germinal role as an antecedent factor to demonstrations of integrity and ability. In other words, benevolence comes first. Fundamentally, transparent processes in media architecture, particularly applications that deploy AI and automated decision-making systems, demonstrate benevolence in three primary ways:

- Transparency practices that expose the governance process of how digital technologies and AI in city making are designed and deployed through adopting a position of *mutual vulnerability*.
- Transparency practices that allow for clear participatory placemaking through *consensus formation*. In this respect, the use of AI and automated decision-making systems are moving backwards in signalling benevolence as even though 'people contribute much of the data that algorithmic systems operate on' [5], the lack of transparency prevents clear understanding of the nature of such participation.
- Transparency practices that create *seams* in decision-making processes, particularly as digital technologies and AI are introduced as a way of automating these decision-making processes in the urban context. These seams create space for developing selfhood and enable the resident to exercise their digital right to the city and contribute to ever-changing iterations of value consensus.

While transparency is a value that appears to emerge from value consensus and leads to demonstrations of integrity, it also permits selfhood development and the exercise of a right to influence policy decision-making. These processes are thus antecedent to value consensus, and the visibility of these processes conveys benevolence. It is from this specific form of transparency as benevolence that genuine value articulation and consensus emerge.

6.1 Mutual vulnerability as mutual visibility

As accelerating levels of technological 'smartness' imbue the city, the emerging smart city becomes a site where data and automation fuse [78]. Given the importance of data—how it is collected, modelled, analysed, and used to drive decision-making within this automated context, transparency processes are crucial to demonstrating benevolence to the trustor. These transparency practices expose the governance process of how digital technologies and AI in city making are designed and deployed by adopting a position of *mutual vulnerability*. In turn, this exposure enables a trustor to be informed and understand how decision-making will occur, how this decision-making process was designed and developed, and how decision-making is responsive to the changing needs of participants and the city.

Wiethoff and Hussman note that media architecture avails 'a new, smart construction material that can. . . enhance. . . communication and enable a material dialogue between the city and citizens' [79]. As mentioned above, these communication channels in media architecture, where designed and deployed through trustworthy processes, serve as a 'credible and trustworthy. . . channel for civic dialogue' [80]. By ensuring that the design and deployment of these digital technologies for city making are not opaque and hidden behind black box approaches, decision-makers, urban planners

and policy makers are not given the opportunity to hide behind these technologies. Instead, transparency requirements retain the involvement and accountability of the decision-maker.

In the same vein, how AI and other data-focused technologies are deployed in city making, how a resident as data subject is made visible through data collection and analytical practices (or 'algorithmic visibility') needs to be examined. Algorithmic visibility regimes are 'highly dependent upon contexts and complex social, technical and political arrangements' [81]. These contextual influences on algorithmic visibility articulate *visual regimes* that push the levers of what is made visible and how it is made visible. In other words, these regimes exercise Foucauldian power over who is forced to be visible in a process of transparency. But while these *possibilities* and *representations* of visibility are at once an exercise of power, they also offer a means of answering the normative questions of how such visibility of the data subject should be managed.

Further, as these technologies veer towards prioritising seamlessness in the dataveillance-to-decision process, this automation process may lock in a power imbalance in the visibility–invisibility dichotomy that hides automated decision-making processes behind a black box [82]. This paper argues that a key means of examining and answering these normative questions of algorithmic visibility is through understanding how transparency—that is, the duality in the exercise of power to render the data subject and to decide how much of the processes of data collection and analysis are made visible—demonstrates trustworthiness through mutual vulnerability.

6.2 Consensus formation

While the increasing datafication and deployment of technology 'has the potential to alienate and disenfranchise citizens' [83], Wang and Burdon contend that benevolence arises from the exercise of a resident's digital 'right to the city'.

> In the context of smart cities, this right to the city has been described as a bundle of rights— including personal access to information and free participation in cultural and democratic activities—that takes expression from 'fundamental principles of justice'. This reflects an emerging digital right that gives [residents] the ability to engage in 'participatory city making' and an 'enlarged . . . Lefebvrian "right to the city" that contributes to the 'democratizing [of] cities and their decision making processes' [82].

This exercise of a resident's digital right to the city emerges through benevolent transparency as socio-political and economic forces are permitted to interact and to agitate, ventilate and bring about consensus in values. Media architecture serves as a crucial conduit in the creation of open public spaces to influence the perceived trustworthiness of these digital technologies and AI used in city making.

In particular, city-making as a multilevel exercise in relationship-building inherently requires value consensus and transparency to be part of the decision-making process. In doing so, there are clear benevolent signals conveyed: residents are given the ability to be 'producers of their own wellbeing by having a say and determining the features of their smart community/city' [10]. These trustworthy signals of benevolence then enable ephemeral social groupings form stable collectives that are mutually vulnerable and thus work to establish and uphold shared values [10]. Transparency practices thus allow for clear participatory placemaking through consensus formation. In this respect, the fields of urban interaction design, digital placemaking and media architecture, outperform other domains of expertise that touch on policy-making, regulation and governance in the city.

Lara et al suggest that a smart city is made human-centered by embedding and using digital technologies to allow for 'participatory governance based on the engagement of civil society in the processes of urban transformation' [10]. However, they stress that such participatory governance and engagement with urban transformation may need to go beyond merely engaging 'residents in the process of building the vision [of the city] for the future' [10], or an ubiquitous approach that recognises 'all passers-by. . . [as] stakeholders whose needs should influence design decisions' [83]. Instead, they suggest that these consensus building interventions also include 'helping in [city] building itself, through co-design and public-private-academia-community partnerships—so called quadruple helix model partnership' [10].

Likewise, Huh et al suggests that designers of such technologies design to anticipate and embed 'social, situational, cultural, and other contextual factors into account' [84], which Foth et al articulate as 'citizen-ability'— the use of design thinking and human-computing interface techniques to create and prioritise engagement with 'new polities and civics' [83]. In this way, the city becomes the interface to allow consensus building to occur, and media architecture the means to 'resurrect the significance and use of town halls, civic squares and public spaces of the city' [83]. These include the 2012 and 2014 projects of *Discussions in Space* at Federation Square, Melbourne, and *Mégaphone* in Montreal.

The notion of consensus is consequently a powerful infrastructural antecedent of democratic data governance. In doing so, these technologies for consensus building resist the 'top-down design of the system [that] fails to capture the complexity of real life at the granular street level of cities' and where 'access controls replicate existing inequalities in the city at a digital level, creating "splintering urbanism"' [9]. However, beyond consensus about the future design of cities, benevolence in the design and deployment of media architecture technologies require us to come together to build consensus about how data collection and use, how algorithmic decision-making and AI is used benevolently in the city. Jobin et al point to over sixty guidelines aimed at establishing TAI developmental and deployment processes [85]. These guidelines take an interdisciplinary approach to increasing the development of TAI, focusing on making AI more explainable [86–89] and more suited to serving our desire for greater equality and distribution of benefits. In response to this process of consensus formation, Thiebes et al articulate five TAI principles that may assist in the future development of trustworthy digital technologies for city making: beneficence, non-maleficence, autonomy, justice, and explicability. There is a degree of overlap between the principles.

- *Beneficence* describes scenarios where technology developers and powerholders deploying AI focus on enhancing human

well-being, including advancing human rights and the environment in which human life occurs. Thiebes et al observe that while certain guidelines such as the UK AI Code only focus on human subjects of AI, other guidelines such as the OECD Principles on AI and the EU TAI Guidelines extend beneficence to nature and climate resilience, and even to enhancing economic life. Beneficence is related to the fields of ethical computing and AI ethics, which seek the '[promotion of] wellbeing into AI at the design and development stages' [76].

- *Non-maleficence* describes the avoidance of harm. Thiebes et al suggest that as misuse of data leads to harm, non-maleficence is specifically oriented towards protecting information privacy, which all TAI guidelines orient towards. For media architecture, this principle is critical in its potential impact on urban life and requires technology developers and powerholders to 'sincerely adhere to ethical and other predefined principles' and for AI systems to act 'honestly and consistently' [76].
- *Autonomy* refers to a general promotion of 'human autonomy, agency, and oversight' where humans are given the ability to 'decide at any given time' [76, 90]. Thiebes et al suggest that autonomy relates to a form of 'openness', where the system is designed to be able to 'give and receive ideas' as a means of communicating trustworthiness [91, 92]. However, they observe that this particular principle is not unanimously captured as a key aspect within the selection of TAI guidelines studied. Instead, for the guidelines that do not accept human autonomy as a key principle, human autonomy is balanced against the other technical goals of the AI system [76]. In contrast, there is a clear strain in media architecture discourse that seeks civic participation in co-designing the city [37], [93].
- *Justice* describes a spectrum of ethical goals set out within each of the TAI guidelines. Thiebes et al note that these can take many forms, including the development and deployment of AIs to address and 'amend past iniquities like discrimination'; more equitable distribution of benefits; or the prevention of new harms. Justice considerations aim at removing bias or 'quantifying the fairness or absence thereof in AI-based systems' [76]. Similarly, Foth describes 'opportunities to explore tinkering with algorithmic might to bring about a diversity dividend and increased innovation capacity in cities [5].
- *Explicability* describes two requirements: the ability for AI to be explained and understood by human users and human subjects and the ability for AI (and the technology developers and powerholders) to be held accountable for predictions and decision outputs. However, Thiebes et al, who observed that within the TAI guidelines studied the degree of explicability varied, generally adopt the term 'transparency' for the mechanism by which TAI is realised. This principle of explicability is the 'most prevalent theme in contemporary AI research' as a reaction to the opaque nature of AI-based systems and how these systems are 'often inaccessible and non-transparent to humans' [76].

6.3 Seamfulness to allow for experiential human moments in decision-making

A 'seam' is defined as a hold point in an automated decision-making process that acts as a 'stopgap that reduces the negative impact' of the relevant digital technology on individual residents [82]. These hold points permit human intervention in the automated process and the exercise of human autonomy. In certain scenarios, these hold points could provide access to the types of data collected and used, as well as the analytical frameworks applied to analyse the collected data, in order to enable such intervention and exercise of autonomy.

Transparency practices create seams in decision-making processes, particularly as digital technologies and AI are introduced as a way of automating decision-making processes in the urban context. However, there is often a technological optimism that comes with digital technologies and AI and their ability to improve city making. This focus on intensifying the use of digital technologies, AI and other open governance processes to achieve a smarter city can neglect the human element subjected to such data collection and decision-making processes. Instead, these seams resist problematic seamless approaches to data collection-to-decision output processes, and hold space for developing selfhood and enabling the resident to exercise their digital right to the city and contribute to ever-changing iterations of value consensus. Fundamentally, these seams should also allow trustors to experience and provide feedback based on these decision-making processes. This allows us to embed thoughtfulness in our experience of transparency and the use of digital technologies and AI. One example in which seamfulness can play out is in the adoption of mediation theory where Verbeek's hybridity is adopted as a means of using thoughtful seams in the design and deployment of digital city making technologies to provide incremental 'range [in] functionality, without being locked in' to extreme responses based on prioritising seamlessness [83].

This paper suggests that seams play an important role particularly where data is collected, used and disclosed. These seams provide the means to experience such data collection practices in a visceral way through hold points in the decision-making process [94]. While digital technology and AI can be used to build and achieve value consensus between urban planners, policy makers and residents, it can also be oriented at a personal level to better the individual. For example, Calo observes that if we shift our approach from using digital technologies under '"command-and-control" regulations' to adopt innovative ways of experiencing data and digital technologies used in city making, there may be opportunities to 'nudge' residents to better understand and safeguard transparency practices. In this case, Calo suggests the traditional regulatory strategy of privacy notices that transparently reveal data collection-to-decision output processes does not take in other benevolent and innovative approaches to putting a data subject in the right model mindset to consider how their data is being used [94].

Benevolence is thus framed as a positive orientation towards the individual within the dataveillance forces of automated technologies. This positive orientation seeks to better individual residents to help them consider how their data is being collected and used. However, Lara et al observe that there is a need to embrace and embed

these technologies to allow for individual fulfilment and wellbeing [10]. To achieve wellbeing and a greater sense of community in our cities, they suggest we require a form of 'neighbourhood social capital' that is based in, among other things, norms of reciprocity—akin to this paper's notion of mutual vulnerability; civic participation—again, the processes that seek and support consensus formation; and 'trust in others, and the benefits of membership' [10]. That is, the ability to benefit individually from being able to intervene creates and maintains the human-oriented seam in the wider joint activity of city building.

This positive orientation towards the individual necessitates protecting spaces for selfhood development through articulating the technological processes—decision-making or otherwise—that operate on the data collected about the individual resident. The articulation of these processes, particularly through information privacy regulation, produces 'seamful stopgaps' within the data-extraction processes, giving space to the resident to 'undertake activities of self-definition and understanding' [82]. In doing so, these seams create spaces for selfhood and exploration of the digital right to the city [82].

7 CONCLUSION

As media architecture technologies increasingly use data-driven technologies and AI to understand and shape the practice of city making, there is a proportionate need to prioritise transparency practices that enable signals of benevolence to build trusted relationships between policy makers, urban designers and residents. Transparency practices should be built to reveal mutual vulnerability by ensuring that what data is collected, stored and used to drive decision-making is not asymmetrically oriented on residents as data subjects. Instead, these transparency practices should require decision-making processes to be revealed. In this way, there is a mutual vulnerability in mutual visibility that enables a resident trustor to see how the decision-making process around city making is using their collected data.

In turn, this mutual vulnerability ensures that the resultant scope, development, use, and deployment of these automated decision-making technologies are not operating in theoretical vacuums behind black boxes but are deeply embedded in the trust relations of those cities that allow active and informed civic participation. As such, this specific form of transparency surrenders access asymmetry to ever-increasing and complex datasets to consider what is responsible data practice. Such surrender of access asymmetry can thus emerge as a new way of considering how automated decision-making systems can be designed and deployed in a trustworthy way to affect how we live, work, and play in these sensorised urban environments.

The emergent form of transparency enables residents to decide how they will be surveilled, and how they might participate in the processes of value consensus and use of their digital right to the city. Where transparent practices expose mutual vulnerability through visibility of both data handler and data subject, boundary-management strategies of seamfulness can be adopted to carve out spaces for autonomy, dignity and selfhood development.

REFERENCES

[1] Sarah Widmer. 2016. Experiencing a Personalised Augmented Reality: Users of Foursquare in Urban Space. In Louise Amoore and Volha Piotukh (Eds.) Algorithmic Life: Calculative Devices in the Age of Big Data. Routledge, London. 57-71.

[2] Tan Yigitcanlar, Nayomi Kankanamge, Massimo Regona, Andres Ruiz Maldonado, Bridget Rowan, Alex Ryu, Kevin C Desouza, Juan M Corchado, Rashid Mehmood and Rita Yi Man Li. 2020. Artificial Intelligence Technologies and Related Urban Planning and Development Concepts: How Are They Perceived and Utilized in Australia? J. Open Innov. Technol. Mark. Complex. 6, 187.

[3] Martin Brynskov, Peter Dalsgaard, and Kim Halskov. 2013. Understanding media architecture (better): one space, three cases. In Proceedings of the workshop on interactive city lighting (CHI 2013). ACM, New York.

[4] Martyn Dade-Robertson. 2013. Architectural User Interfaces: Themes, Trends and Directions in the Evolution of Architectural Design and Human Computer Interaction. International Journal of Architectural Computing 11(1). 1-20.

[5] Marcus Foth. 2017. The Software-Sorted City: Big Data & Algorithms. In Digital Cities Workshop. 26 June.

[6] Ian Mcarthur and Fang Xu. 2021. metaPLACE: Co-designing Sino-Australian Urban Media for Participatory Placemaking. MAB20. 28 June - 2 July, Amsterdam and Utrecht, Netherlands.

[7] Joel Fredericks, Luke Hespanhol and Martin Tomitsch. 2016. Not Just Pretty Lights: Using Digital Technologies to Inform City Making. MAB, 1 – 4 June, Sydney, Australia.

[8] Andrew Vande Moere and Dan Hill. 2012. Designing for the Situated and Public Visualization of Urban Data. J Urban Technology. 19, 2, 25-46.

[9] Brydon Wang. 2021. The Seductive Smart City and the Benevolent Role of Transparency. Interaction Design and Architecture(s) Journal 48, 100-121.

[10] Alexander Prado Lara, Eduardo Moreira Da Costa, Thiago Zilinscki Furlani and Tan Yigitcanlar. 2016. Smartness that matters: towards a comprehensive and human-centred characterization of smart cities. J. Open Innov. Technol. Mark. Complex 2, 8.

[11] Gleb Beliakov. 2007. Construction of Aggregation Operators for Automated Decision Making via Optimal Interpolation and Global Optimization. J, Industrial and Management Optimization, 3, 2, 193–208.

[12] Alessandro Aurigi, Katharine Willis and Lorena Melgaco. 2016. From 'Digital' to 'Smart': Upgrading the City. In Proceedings of the 3rd Conference on Media Architecture Biennale (MAB). ACM, New York. 10, 1–4.

[13] Marcus Foth and Glenda Amayo Caldwell. 2018. More-than-Human Media Architecture. In Proceedings of the 4th Media Architecture Biennale Conference (MAB18).

[14] Malik Tubaishat and Sanjay K Madria. 2003. Sensor Networks: An Overview. IEEE Potentials 22, 2, 20.

[15] M Mazhar Rathore, Anand Paul, Won-Hwa Hong, HyunCheol Seo, Imtiaz Awan and Sharjil Saeed. 2018. Exploiting IoT and Big Data Analytics: Defining Smart Digital City Using Real-time Urban Data. Sustainable Cities and Society, 40 (July 2018), 600-610.

[16] Nivan Ferreira, Jorge Poco, Huy T Vo, Juliana Freire and Claudio T Silva. 2013. Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips. IEEE Transactions on Visualization and Computer Graphics, 19, 12 (December 2013), 2149-2158.

[17] Fabio Leccese, Marco Cagnetti and Daniele Trinca. 2014. A Smart City Application: A Fully Controlled Street Lighting Isle Based on Raspberry-Pi Card, a ZigBee Sensor Network and WiMax. Sensors, 14 (December 2014), 24408-24424.

[18] Jascha Franklin-Hodge. 2019. Foreword. In B Green. The Smart Enough City: Putting Technology in Its Place to Reclaim Our Urban Future, MIT Press.

[19] OECD. 2015. Emerging Issues: The Internet of Things, OECD Publishing.

[20] Mark Burdon. 2020. Digital Data Collection and Information Privacy Law. Cambridge University Press, UK.

[21] Rob Kitchin. 2016. The Ethics of Smart Cities and Urban Science. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374, 2083 (December 2016) 20160115.

[22] Patrick Sisson. 2018. Your City Is Watching You: How Machine Learning and "Computer Vision" Will Transform Our Cities. Curbed (17 January 2018). https://www.curbed.com/2018/1/17/16897222/machine-learning-urban-planning-sidewalk-labs

[23] Esther Keymolen and Astrid Voorwinden. 2020. Can We Negotiate? Trust and the Rule of Law in the Smart City Paradigm. International Review of Law, Computers & Technology, 34, 3. 233-253.

[24] Brydon Wang. 2021. The Machine Metropolis: Introduction to the Automated City. In Brydon T Wang and CM Wang (Eds), Automating Cities: Design, Construction, Operation and Future Impact (Advances in 21st Century Human Settlements), Springer.

[25] J Scott, (1998) Seeing Like a State. How Certain Schemes to Improve the Human Condition Have Failed, Yale University Press.

[26] Alessandra Pieroni, Noemi Scarpato, Luca Di Nunzio, Francesca Fallucchi, Mario Raso. 2018. Smarter City: Smart Energy Grid Based on Blockchain Technology. Int. J. Advanced Science Engineering Info. Technology, 8. 298-306.

[27] Mark Andrejevic and Mark Burdon. 2015. Defining the Sensor Society. Television & New Media, 16, 1. 19-36.

[28] Jennifer Stein and Scott Fisher. 2019. Smart Citizens, Participatory Urbanism, and the Future of City Design. In Tridib Banerjee and Anastasia Loukaitou-Sideris (Eds.) The New Companion to Urban Design. Routledge, London.

[29] Dan Hill. 2020. Small Pieces Loosely Joined: Practices for Super-Local Participative Urbanism. Architectural Design, 90, 3 (May 2020), 66-71.

[30] Mark Austin, Parastoo Delgoshaei, Maria Coelho and Mohammad Heidarinejad. 2020. Architecting Smart City Digital Twins: Combined Semantic Model and Machine Learning Approach. Journal of Management in Engineering, 36, 4, 04020026.

[31] Z Gao, Y Song, C Li, F Zeng and F Wang. 2017. Research on the Application of Rapid Surveying and Mapping for Large Scale Topographic Map by UAV Aerial Photography System. Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2/W6, International Conference on Unmanned Aerial Vehicles in Geomatics, 4-7 September 2017, Bonn, Germany, 121–125.

[32] Tianhu Deng, Keren Zhang, Zuo-Jun Shen. 2021. A Systematic Review of a Digital Twin City: A New Pattern of Urban Governance toward Smart Cities. J. Management Science and Engineering, 6, 125–134.

[33] R El Meouche, I Hijazi, P A Poncet, M Abunemeh and M Rezoug. 2016. UAV Photogrammetry Implementation to Enhance Land Surveying, Comparisons and Possibilities. Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W2 11th 3D Geoinfo Conference, 20-21 October 2016, Athens, Greece, 107-114.

[34] Liang Li, Xingyan Cao, Qiang He, Jie Sun, Bingsong Jia and Xiang Dong. 2019. A New 3D Laser-Scanning and GPS Combined Measurement System. Comptes Rendus Geoscience, 351, 7 (October 2019), 508-516.

[35] Shaohua Wang, Yeran Sun, Yinle Sun, Yong Guan, Zhenhua Feng, Hao Lu, Wenwen Cai and Liang Long. 2019. A Hybrid Framework for High- Performance Modeling of Three-Dimensional Pipe Networks. Int. Journal of Geo-Information, 88, 441.

[36] Anna Luusua, Henrika Pihlajaniemi and Johanna Ylipulli. 2016. Northern Urban Lights: Emplaced Experiences of Urban Lighting as Digital Augmentation. In Nicholas S Dalton, Holger Schnädelbach, Mikael Wiberg, and Tasos Varoudis (Eds) Architecture and InteractionL Human Computer Interaction in Space and Place, Springer, Cham, 275-297.

[37] Irina Paraschivoiu, Janset Shawash, Marta Dziabiola, Narmeen Marji, Alexander Meschtscherjakov, and Mattia Thibault. 2022. Co-designing the Affective City: Speculative Explorations of Affective Place-Based Experiences. In Proceedings of the 25th International Academic Mindtrek Conference (Academic Mindtrek '22). ACM, New York. 313–317.

[38] Bernd Resch, Anja Summa, Günther Sagl, Peter Zeile, and Jan-Philipp Exner. 2015. Urban Emotions—Geo-Semantic Emotion Extraction from Technical Sensors, Human Sensors and Crowdsourced Data. In Georg Gartner and Haosheng Huang (Eds) Progress in Location-Based Services. Springer, Cham. 199–212.

[39] David Lehr and Paul Ohm. 2017. Play with the Data: What Legal Scholars Should Learn about Machine Learning. UC Davis Law Review, 51, 653-718.

[40] Ziad Obermeyer, Brian Powers, Christine Vogeli and Sendhil Mullainathan. 2019. Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations. Science, 366, 6464 (October 2019), 447-453.

[41] Toon Calders and Indre Zliobaite. 2014. Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures. In Bart Custers, Toon Calders, Bart Schermer, Tal Zarsky (Eds), Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases. Springer, 43–57.

[42] Zach Melzer. 2021. Media Architecture as Privatization: Rethinking the Role of Space in the Media Architecture of Yonge-Dundas Square. MAB20, June 28- July 02, Amsterdam and Utrecht, Netherlands.

[43] Robert D Putnam. 1995. Bowling Alone: America's Declining Social Capital. Journal of Democracy, 6. 65-78.

[44] Pamela Paxton. 1999. Is Social Capital Declining in the United States? A Multiple Indicator Assessment. American Journal of Sociology, 105, 1 (July 1999), 88-127.

[45] Petter Bae Brandzaeg, Antoine Pultier and Gro Mette Moen. 2019. Losing Control to Data-Hungry Apps: A Mixed-Methods Approach to Mobile App Privacy. Social Science Computer Review, 37, 4, 466-488.

[46] Jukka Vuorinen, Aki Koivula and Ilkka Koiranen. 2018. Trust and Privacy in Messaging. IADIS International Journal on WWW/Internet, 16, 1, 56-69.

[47] Tobias Kroll and Stefan Stieglitz. 2021. Digital Nudging and Privacy: Improving Decisions about Self-disclosure in Social Networks. Behaviour & Information Technology, 40, 1, 1-17.

[48] Emmanuel W Ayaburi and Daniel N Treku. 2020. Effect of Penitence on Social Media Trust and Privacy Concerns: The Case of Facebook. International Journal of Information Management, 50, 171-181.

[49] Catholijn M Jonker and Jan Treur. 1999. Formal Analysis of Models for the Dynamics of Trust Based on Experiences. In Proceedings of the European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW). Springer, Lecture Notes in Computer Science, 221-231.

[50] Greg Elofson. 1998. Developing Trust with Intelligent Agents: An Exploratory Study. In Cristiano Castelfranchi and Yao-Hua Tan (Eds), Trust and Deception in Virtual Societies, Springer, 125-138.

[51] Thomas W Simpson. 2012. What Is Trust? Pacific Philosophical Quarterly, 93, 4, 550-569.

[52] Mila Hakanen and Aki Soudunsaari, 2012. Building Trust in High-Performing Teams. Technology Innovation Management Review, 2, 6 (June 2012), 38-41.

[53] Teck-Hua Ho and Keith Weigelt. 2005. Trust Building among Strangers. Management Science, 51, 4 (April 2005), 519-530.

[54] D Harrison McKnight and Norman L Chervany. 2001. Trust and Distrust Definitions: One Bite at a Time. In Rino Falcone, Munindar Singh and Yao–Hua Tan (Eds) Trust in Cyber-Societies: Integrating the Human and Artificial Perspectives, Springer, 27-54.

[55] Roger Mayer, James Davis and David Schoorman. 1995. An Integrative Model of Organizational Trust. The Academy of Management Review, 20, 3, 709-734.

[56] Maria Baghamian. 2020. Foreword. In Judith Simon (Ed), The Routledge Handbook of Trust and Philosophy, Routledge.

[57] Bernard Barber. 1983. The Logic and Limits of Trust, New Rutgers University Press.

[58] David G Carneval and Barton Wechsler. 1993. Trust in the Public Sector: Individual and Organizational Determinants. Administration & Society, 23, 4, 471-494.

[59] Julian B Rotter. 1980. Interpersonal Trust, Trustworthiness, and Gullibility. American Psychologist, 35, 1, 1-7.

[60] James S Coleman, 1990. Foundations of Social Theory, Harvard University Press.

[61] John K Rempel, John G Holmes and Mark P Zanna. 1985. Trust in Close Relationships. Journal of Personality and Social Psychology, 49, 1. 95- 112.

[62] David Lewis and Andrew Weigert. 1985. Trust as a Social Reality. Social Forces, 63, 4, 967-985.

[63] Gregory A Bigley and Jone L Pearce. 1998. Straining for Shared Meaning in Organizational Science: Problems of Trust and Distrust. Academy of Management Review, 23, 3 (July 1998), 405-421.

[64] F David Schoorman, Roger C Mayer and James H Davis. 2007. An Integrative Model of Organizational Trust: Past, Present, and Future. The Academy of Management Review, 32, 2, 344-354.

[65] Sirkka L Jarvenpaa, Kathleen Knoll and Dorothy E Leidner. 1998. Is Anybody Out There? Antecedents of Trust in Global Virtual Teams. Journal of Management Information Systems. 14, 4 (Spring 1998), 29-64.

[66] Thomas W Lauer and Xiaodong Deng. 2007. Building Online Trust through Privacy Practices. International Journal of Information Security, 6, 323-331.

[67] Helge Svare, Anne Haugen Gausdal and Guido Möllering. 2020. The Function of Ability, Benevolence, and Integrity-Based Trust in Innovation Networks. Industry and Innovation, 27, 6, 585-604.

[68] Hwee Hoon Tan and Augustine KH Lim. 2009. Trust in Coworkers and Trust in Organizations. Journal of Psychology, 143, 1, 45-66.

[69] Graham Dietz and Deanne N Den Hartog. 2006. Measuring Trust Inside Organisations. Personnel Review, 35, 3, 557-588.

[70] Larue T Hosmer. 1995. Trust: The Connecting Link between Organizational Theory and Philosophical Ethics. Academy of Management Review, 20, 2 (April 1995) 379-403.

[71] Daniel J McAllister. 1995. Affect- and Cognition-Based Trust as Foundations for Interpersonal Cooperation in Organizations. Academy of Management Journal, 38, 1 (February 1995), 24-59.

[72] Roderick M Kramer and Tom R Tyler (Eds). 1996. Trust in Organizations: Frontiers of Theory and Research, Sage.

[73] Mark E Warren (Ed). 1999. Democracy and Trust, Cambridge University Press.

[74] D Harrison McKnight, Michelle Carter, Jason Bennett Thatcher and Paul F Clay. 2011. Trust in a Specific Technology: An Investigation of its Components and Measures. ACM Transactions on Management Information Systems, 2, 2 (July 2011), 12.

[75] Jonathan W Palmer, Joseph P Bailey and Samer Faraj. 2000. The Role of Intermediaries in the Development of Trust on the WWW: The Use and Prominence of Trusted Third Parties and Privacy Statements. J. Computer-Mediated Communication, 5, 3, JCMC532.

[76] Scott Thiebes, Sebastian Lins and Ali Sunyaev. 2021. Trustworthy Artificial Intelligence. Electronic Markets, 31, 447-464.

[77] Michael Lewis, Katia Sycara and Phillip Walker. 2018. The Role of Trust in Human-Robot Interaction. In Hussein Abbass, Jason Scholz and Darryn Reid (Eds), Foundations of Trusted Autonomy, Springer, 135-159.

[78] Marcus Foth, Irina Anastasiu, Monique Mann, and Peta Mitchell. 2021. From Automation to Autonomy: Technological Sovereignty for Better Data Care in Smart Cities. In Brydon T Wang and CM Wang (Eds), Automating Cities: Design, Construction, Operation and Future Impact, Springer, Advances in 21st Century Human Settlements, 319-343.

[79] Alexander Wiethoff and Heinrich Hußmann. 2017. Introduction. In Alexander Wiethoff and Heinrich Hußmann (Eds), Media Architecture: Using Information and Media as Construction Material, Degruyter, Berlin, 1-2.

[80] Joel Fredericks, Luke Hespanhol and Martin Tomitsch. 2016. Not Just Pretty Lights: Using Digital Technologies to Inform City Making. MAB, 1 – 4 June, Sydney, Australia.

[81] Andrea M Brighenti. 2010. Visibility in Social Theory and Social Research, London: Palgrave Macmillan.

[82] Brydon T Wang and Mark Burdon. 2021. Automating Trustworthiness in Digital Twins. In Brydon T Wang and CM Wang (Eds), Automating Cities: Design, Construction, Operation and Future Impact, Springer, Advances in 21st Century Human Settlements, 345-365.
[83] Marcus Foth, Martin Tomitsch, Christine Satchell, M. Hank Haeusler. 2015. From Users to Citizens: Some Thoughts on Designing for Polity and Civics. In OzCHI'15: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction, December 2015, 612-633.
[84] Jina Huh, Mark S Ackerman, Thomas Erickson, Steve Harrison and Phoebe Sengers. 2007. Beyond usability: taking social, situational, cultural, and other contextual factors into account. In Proceedings of CHI'07 Extended Abstracts on Human Factors in Computing Systems, San Jose, CA, 28 April – 3 May.
[85] Anna Jobin, Marcello Ienca and Effy Vayena. 2019. The Global Landscape of AI Ethics Guidelines. Nature Machine Intelligence, 1, 389-399.
[86] EU High-Level Expert Group on AI. 2019. EU Ethics Guidelines for Trustworthy AI, 8 April 2019.
[87] OECD. 2019. Principles of AI, May 2019.
[88] Asilomar AI Principles. 2017. Asilomar Conference on Beneficial AI, 5-8 January, Asilomar, CA.
[89] Department of Health and Social Care. 2019. UK AI Code.
[90] Luciano Floridi, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Fracesca Rossi, Burkhard Schafer, Peggy Valcke and Effy Vayena. 2018. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28, 4, 689-707.
[91] Aneil K Mishra. 1992. Organizational Responses to Crisis: The Role of Mutual Trust and Top Management Teams. PhD Thesis, University of Michigan.
[92] Paul L Schindler and Cher C Thomas. 1993. The Structure of Interpersonal Trust in the Workplace. Psychological Reports, 73, 2, 563-573.
[93] Marcus B Lane. 2005. Public Participation in Planning: an intellectual history. Australian Geographer, 36, 3, 283–299.
[94] M Ryan Calo. 2012. Against Notice Skepticism in Privacy (and Elsewhere). Notre Dame Law Review, 87, 3, 1027-1072.