Ratio Working Paper No. 355:

The artificial intelligence (AI) data access regime: what are the factors affecting the access and sharing of industrial AI data?

Per-Olof Bjuggren & Vicky Long

RATIO Näringslivets forskningsinstitut

Box 3203
103 34 Stockholm

Besöksadress:
Sveavägen 59, 4 tr

The artificial intelligence (AI) data access regime: what are the factors affecting the access and sharing of industrial AI data?[1]

Per-Olof Bjuggren

per-olof.bjuggren@ratio.se; The Ratio Institute

per-olof.bjuggren@ju.se; Jönköping University

Vicky Long

vicky.long@hh.se; Halmstad University

vicky.long@ratio.se; The Ratio Institute

Abstract:

This paper decomposes the factors that govern the access and sharing of machine-generated industrial data in the artificial intelligence era. Through a mapping of the key technological, institutional, and firm-level factors that affect the choice of governance structures, this study provides a synthesised view of AI data-sharing and coordination mechanisms. The question to be asked here is whether the hitherto *de facto* control—bilateral contracts and technical solution-dominating industrial practices in data sharing—can handle the long-run exchange needs or not.

Key words: Artificial intelligence (AI), governance structure, intellectual property rights (IPRs), data trade, industrial data

JEL Classification: D23, K1, K24, L14, L86, O3

[1] This draft paper was presented at the Nordic Corporate Governance Network workshop (28 June 2021), organised by Copenhagen Business School (CBS).

Table of contents

1 Introduction

In this paper, we take a closer look on how machine-generated data are accessed and shared in the new economy of artificial intelligence (AI) and what the consequences are for economic organisation.

We are entering an AI era. The internet of things is already here. Billions of things—sensors, actuators, drones, and robots—are connected. There is an explosion of things-generated data, on which business, engineering and financial processes can draw. Applications are exemplified in manufacturing robots, automated financial investing, and disease mapping. Machine learning algorithms thrive on access to big and varied datasets. Phrases like “it’s all about data” and “data is oil” speak of the importance of data access and data trade for economic efficiency (cf. OECD, 2015).

Theoretically, knowledge exchange is hardly new. Sharing and coordination of decentralised knowledge has always seen as essential for an economy to function and prosper (Hayek, 1945). What is new, in the era of AI, is the uncertainties in specifying the transaction costs involved in data trade.

The complexities related to AI data trade illustrated are:

Firstly, and on the technological front, data are a moving target, highly dynamic in *volume*, *variety,* and *velocity* (Gandomi and Haider, 2014). This also goes for the actors—enterprises, machines, and consumers—who create that data. That causes problems, not least in allocating access and control, in a networked environment.

Secondly, and at the institutional front, legal analyses show that no legal scheme (yet) provides a framework for exclusive rights in machine-generated data; for example, copyright only protects human creativity (Wiebe, 2017). The introduction of new property rights for machine-generated data is argued as neither necessary nor economically justified for the actors involved (Drexl, 2017). Current data-access practices are heavily dominated by bilateral contracts and technical protection means (Martens, 2018); while the former has limits in up-scaling,[2] the latter introduces factual exclusivity. This seems to be against the need for value creation from AI data, which requires channelling as many contractual parties as possible (in, e.g., the internet of things).

Thirdly, and at the firm level, AI strategy requires a data strategy: (a) the value creation processes are interrelated. Many firms are not only producers of data, but they also rely on access to others’ data at the same time. There is a high degree of interdependence between exchange partners in value creation. Often the value of the data can only be realised if all parties collaborate (in, e.g., autonomous driving); and (b) there is a large upfront cost in collecting, sorting, storing, and transmitting data, and in training algorithms, putting pressures on access (rights) to data (Martens, 2018). Information is nonrival: one person´s use of information does not reduce or diminish another person´s use (Arrow, 1962); so is

[2] This refers to the problem of a small numbers (of dealers) in transactions discussed by Williamson (1975) when he uses the number of buyers or sellers to define the structure of an exchange relationship.

machine-generated AI data (Varian, 2019). This makes the positive-sum game argument for collaboration even stronger here.[3]

This study represents an attempt to decompose the factors that govern the access to and sharing of machine-generated industrial data in the AI era. To accomplish this, some fundamental factors have to be considered. These factors are of a technical as well as an institutional nature, and they provide a framework within which firms can act as they strive to take advantage of the new techniques offered by AI and machine learning (ML). Text analyses on the doctrines and laws related to data protection are primarily used. Firm-level data were mainly collected through a pilot study, with interviews and discussions with a firm engaged in computer vision and smart city solutions.

The focus is largely, but not exclusively, on machine-generated business-to-business (B2B) data. This is because the institutional frameworks related to B2B data and B2C (business-to-consumer) data are different. B2B data are sometimes referred to as industrial data, in contrast to personal (B2C) data.[4] In the following text, we use the word data instead of industrial data or B2B data for the sake of simplicity. Moreover, the labels data trade, data sharing, access to data and data coordination are used interchangeably in this paper, denoting inter-firm data transfers.

This study contributes a synthesised view on AI data-sharing practices and coordination mechanisms. It also contributes to a general understanding of how the boundaries and interdependence of firms change over time and the impact of such changes on economic transformation.

Data coordination is subject to both technical and human-devised constraints. This paper is therefore structured as follows: in Section 2, we discuss the technological nature of AI data (i.e., how data are produced); in Section 3 we discuss the formal and informal rules within which access to and transactions of data are coordinated and organised. We delineate the institutional problems encountered in AI data sharing. While these two types of constraints form the background to the choice of governance structures available to the actors in the new economy, in Section 4 we look closer at key governance issues and AI data asset specificity. Section 5 examines firm-level appropriability hazards. Section 6 concludes the paper. In this way, we put the transaction cost economising into a larger economic framework and discuss the relevant trade-offs.

2 Product characteristics: what is AI data?

Machine-generated data is ubiquitous and access exclusivity can be achieved either by *de jure* rights (legal restrictions) or *de facto* control (e.g., technical means). While the legal aspect is discussed in Section 3, in this section, we discuss the (technical) traits of the AI data that are subject to coordination.

---

[3] Data being public goods (see the 2015 OECD *Data-Driven Innovation Report*: the more sharing of data, the better for society).

[4] Personal data, as defined by the GDPR, are data that can identify a person, directly or indirectly. B2B data here, may be crudely referred to as anonymised data in which personal information cannot be identified and from which it cannot be derived (cf. Schneier, 2007). For example, billions of anonymised x-ray bone-defect photos can be used in training ML algorithms.

2.1 Types of AI data

Machine-generated data are heterogenous in feature, function and benefit, both alone and in combination.

Firstly, there are *unstructured versus structured* data, synonymous with *unlabelled versus labelled* data as discussed in the informatics engineering field. The former refers to raw data, bits and bytes data (cf. Li et al., 2020) which, according to Gandomi and Haider (2015), accounts for 95% of big data. Unstructured data often have problems with accuracy, completeness and interoperability (Edinburgh Report, 2020), and they are not subject to direct sharing. There is an upfront cost in sorting and categorising the (unstructured) data.

Secondly, data are dynamically changing and interlinked in value creation. The changes are sometimes described in terms like 3Vs or 4Vs, namely in *volume*, *variety*, *velocity* and *veracity* (Laney, 2001). AI data processing often occurs in a networked environment: features (of data) do not appear *ex ante*, but they are the result of coordination and analysis. There is a complex chain of actors, infrastructures and activities involved in collecting, segmenting, matching and positioning the data. The value-creation processes are interlinked.

Thirdly, data can be heterogenous in value. In autonomous driving, for example, data can be (the precursor of) *information* (e.g., maps), *knowledge* (e.g., driving behaviour) or *property* (e.g., algorithm-based prediction methods[5]). It is not evident who should gain access and when they should gain access, particularly when data include a combination of information from varied public and proprietary domains.

Academically, there is a divide related to the value of machine-generated AI data.

One approach considers data only "the precursor of information, which is the precursor of knowledge", because "data are defined as uninterpreted symbols; information is data with added meaning; and knowledge is the ability to assign meaning to data in order to gain new information" (Stepanov, 2020, p. 67). Machine-generated data often seem homogenous, lack diversity and are only superior in the short run; consequently, they do not lead to radical innovation (Zurth, 2020; Jones and Tonetti, 2020). "We can think of data being stored in bits, information stored in documents, and knowledge stored in humans" (Varian, 2019, p. 404).

Another approach assigns more value to machine-generated AI data. Data today, as argued in the engineering community, are not simply bigger, but have evolved in significant ways, exemplified with the 1980s' emergence of relational databases (facilitating modelling relational algebra) (Information Technology Vocabulary of the ISO/IEC 20546:2019(en)).[6] A similar position on this can also be found in the social sciences community: data(bases) are used as a *tradeable commodity* (Zech, 2016, 2017). Acquiring data implies costs on infrastructures and on managerial and engineering skills. This is the very rationale underlying the European Union (EU) Database Directive (DBD) of 1996,[7] in which the data(base), often

[5] under the *sui generis* right protection due to a "substantial investment in either the obtaining, verification or presentation of the contents" (Art. 7(1) of the Database Directive).
[6] See https://www.iso.org/obp/ui/#iso:std:iso-iec:20546:ed-1:v1:en ISO/IEC 20546:2019(en) Information technology—Big data—Overview and vocabulary.
[7] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L77/20.

involving (substantial) investments, is treated as a possession, tradeable and licensable.[8] Burnstein (2012) argues that data are splitable, and that showing part of the data does not necessarily reveal all the data (which rejects Arrow's information paradox). Data are dynamically evolving; some data are more valuable than others (i.e., beyond providing mere information).

2.2 Characteristics of AI data

In AI data creation, many mechanisms are at work, notably two-sided market network economies in, for example, platform-based businesses (Rysman, 2009; Gawer, 2014; Mansell and Steinmueller, 2020). We discuss a few key mechanisms in the following paragraphs.

Scale is relevant particularly for platform-based AI data. A digital platform corresponds to the old-fashioned market, where suppliers and customer gather to trade with each other. Dubbed two-sided markets and sometimes multi-sided markets, digital platforms, however, differ from traditional markets by acting as matchmakers for two separate groups, one posting for users and another for advertisers (Evans and Schmalensee, 2016).[9]

There is a classic scale advantage on the supply side: the larger the user group, the smaller the per unit/user cost in storing and delivering content. Investing in datasets for training ML algorithms has a high upfront cost (Colangelo and Maggiolino, 2017). However, after the algorithm has been properly trained, the marginal cost to produce an additional dataset is low. There is, in addition, an element of learning by doing (Arrow, 1962), namely improvement occurs over time, which further reduces the unit cost and/or increases the quality/value. This learning by doing, sometimes referred to as an experience curve or a learning curve, is a longstanding feature of human beings. This cumulative production aspect also applies to ML.

There is also a classic scale advantage emanating from the demand side: the more connected users there are, the more attractive the platform becomes, and then even more users want to join the platform. Connectedness is a key word in AI and ML (Porter and Heppelmann, 2014). This demand-induced scale advantage is sometimes referred to as a network economy (Goldfarb and Trefler, 2018). The network effect comes in different varieties. The classic example is the telephone line: the more users, the more telephones are perceived as useful by other users, and then the more likely that they will join too. Sometimes this telephone line example is referred as a *direct network effect*. An indirect network effect refers to the situation in which the scale attraction comes from user volume in adjacent complementary applications. For example, the number of Google Doc users is affected by the number of Gmail users. In the AI data context, both sorts of network effect exist. More users means more data.

For ML, scaled user data is not only desirable, but also a necessity: the larger the datasets, the more accurately the algorithms can make predictions. So, this direct network effect is more straightforward in AI for competitiveness: competition for data is the core. The indirect network effect is important too. Google—with many applications such as Google Maps and

[8] "Substantial investment" is, however, required to qualify for that right. It is also hard to distinguish whether it refers to machine-generated data or human-collected data; mostly probably, the latter.

[9] In the new digital era, platforms offer information advantages. The search costs (contact costs) can be greatly reduced for both suppliers and customers by finding a good match. However, the information asymmetry is not necessarily changed.

Google Docs—is an example of how a network economy provides an important explanation of why a company grows to be dominant in its (portfolio) line of business.

There are also an economies of scope when merging varied datasets (Martens, 2018): the benefits of using the aggregated data are higher than those of using the data separately.[10] By definition, merged datasets lead to greater total demand, or to a reduction in the average cost, than operating each dataset separately. This is the key argument in favour of wider AI data access (than exclusive ownership): a smart city gets smarter when its trash bins (filling level) data[11], traffic data, demographic data (on, e.g., healthcare needs) and public facility data are merged to tailor services to individual needs. More accurate and insightful predictions can be generated if data are merged. For ML training, data variation *per se* is not necessarily an advantage. A statistical estimation model becomes more reliable when the size of the data increases and the variation of the data reduces. Varied datasets need to be in scale as well to ensure robustness of estimation.

Economies of scope explain why data-driven firms are so data-hungry and collect all the data they can get. Economies of scope, however, are unlikely to continue forever, and sometimes they are subject to diminishing returns (Duch-Brown et al., 2017), for example on film selection (Pilaszy and Tikk, 2009).

The combination of economies of scale and a wide scope (e.g., using digital platforms to gather data) can lead to a natural monopoly, often seen in social media and e-commerce platforms (e.g., Facebook, Amazon). There is a strong network effect, a reinforcement cycle, often associated with first mover advantage (see Section 5). It is also related to the concepts of exclusivity, market power and barriers to entry. So far, platform dominance—a natural monopoly—seems to be primarily a problem in the B2C market where personal data is concerned, rather than in the B2B market, in which the internet of things dominates.

2.3 What do these product characteristics imply for data access?

What do these AI data product characteristics imply for data access and data coordination? AI data are heterogenous, dynamically changing and inter-linked in value creation, implying a dynamic shift in product-service domains. This affects critical decisions that have a bearing on the boundaries of organisations, as well as on the control and allocation of resources within organisations.

To take a step back, the advantages of tight data coordination (access and sharing) are mainly derived from AI data characteristics:

- Data is nonrival in consumption, which means the consumption of data by one (person/firm) does not reduce or diminish its use by another (Varian, 2019). This applies at least to data that are merely information. It is therefore arguable that data access is a more appropriate concept than data ownership.

[10] B(d1,d2) > B1(d1) + B2(d2) (Martens and Mueller-Langer, 2018).
[11] through placing Radio Frequency Identification (RFID) tags and internet-connected ultrasound sensors in bins.

- There is a substantial upfront cost in collecting, sorting, categorising and training (particularly unstructured) data. It is beneficial to skip/recoup that cost through sharing data.
- Value creations are interlinked: many firms are not only producers of data, but also rely on access to others' data at the same time. There is a strong interdependence between exchange partners in the pursuit of joint value. Often the value of the data can only be realised if all parties collaborate (in e.g., the internet of things).
- Economies of scope: the value of the aggregated dataset is bigger than the sum of the values of each dataset. It is often a positive sum game to share the (non-rival) data.
- Coordination of (decentralised) knowledge is important for economic efficiency (Hayek, 1945), a macro argument.
- For small- and-medium-sized actors who do not have an established monopoly in the market (e.g., platform owners with first movers advantage and/or network effect), access to others' data is important.

There are two types of coordination uncertainties derived from AI data characteristics. One is within the firm (hierarchy) and the other is in the market (among traders). The first one can be assessed from a resource-based view of the firm. The second one is related to the transaction cost perspective.

To elaborate, the first uncertainty is related to the processes of AI data value creation and rent-seeking. In the internet of things, for example, data can be split into bits and bytes or into packages. Coordination can involve parts of datasets or whole datasets. Firms can take a gradual approach to data trading and disclosure. The high volatility of AI data means high uncertainty in values transferred and in rents allocated: the rent-generating potential is always changing, and it also depends on the scale and scope of the data shared. It is unclear whether (and where) there would be a capability overlap or complementarity, which affects competition and rent-seeking strategies within the firm (hierarchy).

The second uncertainty is related to the behaviour uncertainty of AI data coordination among the exchange partners. Data changes occur in the 4Vs, volume, variety, velocity and veracity. This dynamic shift applies to both upstream data generation and downstream application (appropriability). This is a disturbance—in frequencies and in variances—of the existing exchange/contracting relations. When the (scope of the) technology system is difficult to specify and when the scope of coordination is wide and changing, it causes monitoring difficulties.

Transaction cost theory (e.g., Williamson, 1975) predicts that—in this specific situation—more hierarchical forms of governance—such as firm and/or relational contracting—are preferred over market (exchange). Hierarchical forms of governance allow managers to internalise external variances, including opportunistic behaviours of partners, and to control them (Thomson, 1967). Relational contracts can specify each party's behaviour at a general level, and there is flexibility in modifying, enforcing and monitoring such contracts, while neoclassic contracts can only handle specific contingencies (Macneil, 1978).

There is, however, a puzzle here. Transaction cost theory—in Williamson's (1985) later elaboration—predicts that a preference for more hierarchical forms of governance should become strongly visible when there are also small numbers of traders. This bargaining problem – caused by small numbers of traders - is arguably also visible in contemporary two-sided platforms in which the platform owner dominates (Barach et al.,2018). This, however, is not necessarily the case for B2B data, where the internet of things is the main source of information. A smart city for example, is a system of systems, where energy system data, traffic system data, public facilities (water) system data, transportation system data work together to make it smart. In other words, there are many cloud-based data analytics platforms, rather than a single dominant platform. Small numbers of traders bargaining problem may not exist to the same extent.

3 Institutional frameworks and AI data access

Institutional frameworks determine how well allocation mechanisms work and therefore function as important premises for knowledge sharing, including AI data sharing.

This section provides a brief review of institutions related to AI data. Institutions here are broadly defined as embracing both formal and informal rules: laws, customs and politics (cf. North, 1986; Davis and North, 1971, pp. 6-7) The basic institutional framework constitutes, as North (1990, p. 4) put it, "the rules of the game" in which human interaction takes place. These rules give the constraints within which humans act to promote their interests. Formal institutions can be understood as placing political constraints on behaviour and informal institutions, such as norms and customs, as placing private constraints on behaviour. In the following, we go through the formal and informal institutions related to AI data.

3.1 Formal institution: *in rem* rights and *in personam* rights

Our discussion of the formal institution on AI data access starts with a division of rights into rights *in rem* (imposed on all people) and rights *in personam* (imposed on specific persons). They are two types of legal protection with implications for ownership and transferability of possession. Typical *in rem* rights include intellectual property rights (IPRs) and tort rights. This distinction is not static, as Hudson (1899) argued that rights *in personam* can give rise to rights *in rem* so far as "torts are founded on contract" (p. 13).

Legal protection *in rem* is what characterises patents, copyrights, trademarks and company secrets, while protection *in personam* characterises contractual protection. *In rem* is a protection against anyone else. *In personam* protection is used to safeguard bilateral relations. Patents and copyrights give exclusive rights against anyone else, and those rights can be traded on markets. This allows transfers of rights to the users who give the highest values to these rights. Of course, how efficiently the market mediates transfer of these rights depends on the cost of using markets for transactions. Perhaps mergers—vertically or horizontally—are a less costly alternative here. Or perhaps high transaction costs preclude any transfer of rights.

The words property rights and ownership are related (see Hodgson, 2015; Hohfeld, 1913). The word ownership can be used for IPRs (*in rem* rights legally valid against anyone). Furthermore, it is important to recognise that legal ownership protection is often not a must for controlling the use of an asset. In absence of *de jure* rights (i.e., ownership rights), (exclusive) control can be obtained *de facto*. Such *de facto* control of data can be established through technical means (e.g., smart cards, PINs, biometric authentication), or with industry-wide standards. With *de facto* control, firms can use contractual arrangements to provide access to data to others and to charge them for that access. However, the contractual rules agreed upon are only valid for the involved contractual parties (an *in personam* right).

In rem (1): intellectual property rights

The two IPRs *in rem* for data protection are copyright and the *sui generis database protection right* ("the *sui generis right*" hereafter) (Wiebe, 2017; Drexl, 2017; Stepanov, 2020). While copyright protects creativity, the *sui generis* right protects investment.

Legal acknowledgement of private property rights can facilitate an efficient allocation of resources and create incentives for investments, innovation and growth (Coase, 1960; Arrow, 1962; Alchian and Demsetz, 1972). This is as close to a unanimous agreement as the academic research can come.

The central debate here is whether data ownership rights should be granted on machine-generated data or not, namely as a novel legal right or not. It is also related to the question of to what extent the existing legal rights are effective in economic transactions, including data trade. We go through the existing institutional arrangements below.

IPRs like patents, copyrights and trademarks are instituted to provide investment incentives to (re)produce such new knowledge. Without IPRs, knowledge is in the public domain. Just like a public good, it can be consumed by many at the same time. IPRs make knowledge a private good protected by law. As a private good, it can be traded in the market. Beside that, IPRs solve the problem of knowledge baptised as an information paradox (Arrow, 1962). The information paradox basically discusses a disclosure dilemma when the product is intangible information or knowledge: if you want to sell new knowledge to someone, that person is likely to want to know what she/he is buying. But as soon as she/he has been told, there is hardly any incentive to pay. This is also relevant to AI data. IPRs solve the problem of both the exclusivity that characterises a private good and the information problem by requiring description when applying for patents and trademarks (after being granted, information can be disclosed without fear that it will be taken without compensation). Solving these two problems enables market-based exchanges of knowledge.

Saying that, it is important to turn to the fact that current IPR regimes do not grant AI data—machine-generated data—an exclusive right. This is in the core of the big legal debate on whether a new exclusive right for AI data is needed (Wiebe, 2017).

The anti-side—the opponents of an exclusive right—argues:

Firstly, data access overrides data ownership in importance: in the new economy of AI and ML, data are the driving force. AI and ML are dependent on access to big data sources. Data are the new oil in this new phase of industrial development. Many firms are producers of data, but that production often relies on access to the data of other players. Access to data is necessary and important (OECD, 2015).

Particularly for small and medium-sized enterprises, the introduction of new rights may provide more drawbacks than benefits. Many stakeholders, contributing to the same data-based business model, have diverse kinds of interests. Therefore, too high a proportion of the ownership rights might have a negative effect on accessibility and not be needed from a cost-benefit perspective for the majority of the actors involved (see also Drexl, 2017).

Secondly, and related to the well-established argument that IPRs give incentives to productive innovations that Arrow, Teece, Stiglitz and the endogenous growth school (Aghion and Howitt, 1998) put forward, it is questionable whether this nearly unanimously agreed IPRs function is needed or not. The unique features of AI data make it necessary to look at IPRs with new eyeglasses. Machines, robots, sensors and actuators will generate data anyway without exclusive rights at place to catch up with the competition (Weibe, 2017). AI and ML seem to develop rapidly without these legal devices. In addition, it is not just the input of data in the creation process that is the issue. The output here in the new AI era is often also of data that can be further fed into a new round of knowledge creation. This output can be of a character close to that of literary or artistic output protected by copyright laws. The only difference is that a machine rather than a human being is the creator and, as pointed out, machine generated data cannot be copyright protected (so far).

The pro side—the proponents of an exclusive right—argues that AI data can be understood as a goods in the economic sense and an object in the legal sense, with all three components qualifying for a property right: semantic information, syntactic information and structural information. There is a need to have a kind of informational property that benefits the data producers, particularly in contracting relations (Zech, 2016, 2017). It is also argued, from a copyright protection perspective, that AI works—though they are created by machines—are sufficiently original to qualify for copyright protection. Therefore, machine-generated data should be included in the scope of copyright protection (Goold et al., 2020).

For industrial data, copyright is not so important, but the *sui generis right* is. Copyright here is like copyright for literary and artistic work. It is about the author, the creativity, the expression (not the idea). Hence, it is only awarded for human creativity and novelty. Creativity can, for example, be embodied in a unique arrangement of the data in a database (European Commission, 2017, 2018). That, in legal terms, is often phrased as the structure of the data(base), namely the way of organising or transforming the data to make sense in the real world. It is important to clarify that the content as such—or the data itself—does not qualify for copyright protection. This is, to a large extent, due to the collection method in the new era of AI and ML (generated by machines, rather than by humans) (Gervais, 2019). As a consequence, copyright protection is not so useful for industrial (B2B) data.

The *sui generis right*—as a means to protect investment—is relatively more relevant here than copyright. The *sui generis right*, however, only exists in Europe. It dates to 1996, when the UK Copyright and Patents Act, 1988 was amended to introduce the *sui generis* database right: an *in rem* protection for "a substantial investment in either the obtaining, verification and/or presentation of the contents". The objective is to provide incentives for investment in the form of setting up and maintaining databases. It does not protect the (individual) data as such. Only the original database as a collection is protected for a period of 15 years. Furthermore, new data generated from a company's activities are not protected. The parts-whole relationship can be phrased so that if one takes a substantial part of the database out (without compensation), it constitutes an infringement. The data themselves are free, but not in packages when substantial investment is concerned. This is very relevant for the reuse of the data. Drexl (2017) criticises the functioning of this protection in the internet-of-things era as being too static and not up to date to handle the new dynamics and real-time services. Moreover, it seems to concern data collected by humans rather than by machines. Presumably that was not an issue in 1996. Consequently, this *de jure* right does not seem so important in the new era of AI.

Trademark is another *in rem* type of protection of possession. However, is difficult to buy in markets. Trade secrets work in a similar way. To get access to information through theft and former employees can be prohibited in both criminal and civil law (labour law). An alternative is to acquire the company associated with the trademark and the trade secrets.

Trade secret is an additional *in rem* type of protection of possession. The EU Trade Secrets Directive of 2016 can be applicable if three criteria are fulfilled: (a) secrecy in the sense that the information contained in the data is not generally known or readily accessible, (b) it has a commercial value and (c) reasonable steps have been taken by the person in control to keep the data secret. Trade secrets can be protected by criminal and/or labour law (i.e., being part of an employment contract that specifies what is considered a trade secret). However, data produced by machines do NOT meet the secrecy condition, as it can be difficult to show commercial value, and/or it is difficult to pinpoint the person in control when data are generated in networks of firms (Drexl, 2017). This type of protection can be classified as an *in personam* type of protection (Stepanov, 2020) in the case of (employment) contracts for example.

To sum up, *in rem* rights in intellectual property are somewhat vague and perhaps even unimportant in the protection of AI data. It is essentially *de facto* control, rather than ownership rights, that has hitherto been in play.

In rem (2): competition law

The distinction between legal ownership rights (*de jure* rights) and *de facto* control is important again here. *De facto* control can be found in technology standards, and/or enforced via technical (control) means to exclude others from use. With *de facto* control, access can be obtained through bilateral contracts. It is observed that the industry practice in data trading has hitherto used bilateral contracts and technical means to gain control (Martens, 2018), and that both means introduce factual exclusivity (i.e., blocking others' access) (Drexl, 2017). The two methods are particularly problematic when the data are reused or resold.

Data access under competition law has long been debated in academia and among competition policy makers, as has the role of antitrust law in facilitating data sharing. Exceptions for access to data can be found in the so-called essential facilities doctrine (EFD), which is based on the idea that a monopolist firm has a duty to share its facilities with anyone who asks for access (including competitors). The EFD, originally developed by the US courts in the 1980s and later also gaining popularity in the EU, is one of the most controversial antitrust issues debated, as it can act as a counterincentive to investment. Moreover, the antitrust liability derived from EFD only concerns monopolistic situations, not general anti-competitive behaviour (Colangelo and Maggiolino, 2017). That is to say, the scope of regulation here is still not sufficient to ensure the effectiveness of the liberalisation process.

Article 102 of the Treaty of the Functioning of the European Union (TFEU) prohibits abusive conduct by companies that have a dominant position in a market. A dominant position can result from advantage of scale, control due to IPRs, like copyright and patents, and/or having *de facto* control by technical means or combinations of these reasons for dominance. Abusive conduct in case of interest here is if the dominant company with an interest in a downstream stage refuses to give access to indispensable data to a firm operating or that is about to operate in the same downstream area. This is considered a refusal to supply that can be prosecuted by the European Court of Justice (ECJ) (Drexl, 2016);

Two illustrative cases of how the ECJ has applied Article 102 in judgments on dominant position and refusal to deliver are *Magil* and *Microsoft*. The *Magil* case concerned three TV stations in Ireland that refused to grant a copyright for information about their coming TV programs. As a consequence, Magil could not offer TV guides to its customers. The three TV stations were the only source of information, and they were considered *de facto* monopolists.

A more interesting case in how the ECJ has reasoned is *Microsoft v. Commission*. Microsoft was considered to have a dominant position not just because of an intellectual property right but because its Windows operating system had emerged as a *de facto* standard in the market due to network effects. In the *Microsoft* case, the issue was access to the interoperability information necessary to get compatibility with the Windows operating system. Without compatibility, there was not so much of a market for downstream networking products. Microsoft refused to provide the license needed for accessibility. One reason was that Microsoft wanted to develop downstream products itself. In the Microsoft case, four requirements for establishing abuse of dominant position were used: (a) the information (data) is indispensable for business in the downstream market, (b) the refusal excludes effective competition, (c) the refusal prevents the emergence of a new demanded product, and (d) there is no objective justification for refusal. Another case involving Huawei dealt with the refusal to license a standard-essential patent. What can be learned from these cases is that these are examples of how economies of network, *de facto* control and *de jure* control based on patents and copyrights can be used to prove abuse of a dominant position in court and hence to force companies to give access to data/information.

Competition law is to a large extent about open access to data when a company has a dominant position. A quite different strategy is for a company on purpose to choose open access. Given that choice, one might ask what kind of thinking is behind such a choice? What are the advantages and disadvantages? One advantage is that to a large extent, cooperation and sharing of information is important in the new era. An example of this is the open-source operating system that characterises Linux. In contrast to Microsoft, the interoperability information is freely available. The advantage is that the feedback from different users helps to develop the system. A disadvantage is of course the lost licence payment.

In personam rights

This extends the discussion further from intellectual property law to contract law and tort law. Tort law is actually an *in rem* right,[12] but in practice, tort/liability is often related to what is specified in the contract. We therefore discuss them together here.

---

[12] Property rights and liability rules are connected (Alchian and Demsetz, 1972). They are both *in rem* rights. Coase (1960) shows in his seminal article how clearly defined legal property rights (implying clearly defined liability) over allocation of resources can work no matter who has the rights. This is the content of his famous Coase theorem. In light of the fact that there is not much data granted with legal property rights, the issue of liability is of concern. Another important aspect of property rights here is the distinction made by Calabresi and Melamed (1972) between what they called entitlements/property rights protected by property and liability rules. A liability rule implies that harm to a property can be compensated with damages. In Calabresi and Melamed it means, for example, that the state can confiscate a private property/entitlement and pay damages corresponding to the value of the confiscated property. Cooter and Ulen (2008) clarify the concept of property rules in line with

This feedback process and output of a literary/artistic character aggravate not just the ownership problem, but also other legal issues such as liability and contracts, which are often regulated under *in personam* rights. Returning to the theme of this paper, the governance of data access and data sharing, a big question is what rules of social interactions (institutions) are needed to produce a high degree of welfare and economic development in this new technological era?

These rules need to give incentives to produce, transact and protect in a manner that promotes economic efficiency, welfare and development. This has implications for how to organise the economy in markets, and other type of organisations that are not dependent on the price as an allocation mechanism. Let us start with the legal ownership question, which is a cornerstone for the question of the organisation of the economy. The next step is to examine the implications for economic organisation and liability.

Last liability (tort law) is also an issue in the AI era. Both multiple actors and multiple types of actors are involved in, for example, the operation of a smart city. Some are machines and robots, involved in the production of data. Some have network capabilities and are capable of orchestrating the burgeoning digital (eco)systems. Would the car producer, the traffic data producer or the pedestrian map producer be liable to a pedestrian involved in a car accident? Three criteria for liability are harm, cause of harm and breach of duty (Cooter and Ulen, 2008, Ch. 8). These criteria apply to human beings. Harm implies that someone's utility/welfare has been negatively affected due to, e.g., injury or libel. The second criterion is that someone else has caused this harm. If these two criteria are fulfilled, the person responsible can be liable to pay damages that restore welfare if a rule of strict liability is applied. However, sometimes breach of duty is also considered. In that case, the legal situation is that instead of strict liability, some kind of negligence rule is applied. As the rules apply to human beings, the legal situation can be somewhat tricky if several stakeholders as well as ML and AI are involved (European Commission, 2017). For example, if a smart car (self-driving car) is involved in an accident with harms to humans and property, it can be tricky to determine liability. Is the owner of the car, the user of the car, the manufacturer, the sensor company that installed the software or another supplier of data that tells the car how to operate responsible?

3.2 Informal institutions

Norms and private ordering can be added to the list of protection *in personam*. Private ordering represents here a bilateral relationship that can involve differing combinations of enforcement through hostages, norms and law. These are extra-legal devices to control possession. Pure private ordering means that parties agree to police a relationship themselves. Established norms (informal regulation) are chosen instead of government regulation. However, the contractual model, which is chosen most often, is more correctly a quasi-private ordering, as it involves a reliance on contracts that can be enforced by the legal system (Lemley, 1998).

The rationale behind private ordering is that contracts often do not handle accessibility *ex post* as anticipated *ex ante*. Contracts, especially long-term contracts, tend to be incomplete

Calabresi and Melamed. If the rights are protected by liability rules, damage paid for past harm can be used as an alternative to an injunction that forbids harm to the owner in the future.

and imperfect (Williamson, 1985 p. 164). A seminal article by Macaulay (1963), which went through a large number of contracts, found that the small print in contracts was not legally enforced for the simple reason that the parties were concerned about keeping their business relationship. Williamson (1985), referring to an article by Galanter (1981), further analysed contractual relations in a discussion of private ordering and reliance on legal enforcement *ex post*. Private ordering refers to a contractual relationship that *ex ante* is characterised as an awareness of contractual incompleteness and therefore anticipates the need for settling conflicts and tries to craft mechanisms to settle conflicts out of court. Williamson further explains how such mechanisms can be created in a discussion of credible commitments, where credible commitments are different types of losses that will be incurred by both parties if the contractual relationship is broken.

The importance of private ordering contracts in B2B data transactions has been pointed out by Wiebe (2017), among others. So, private ordering, in which parties agree to police a relationship themselves, is a way to govern relations concerning the internet and the new information technology. This can be viewed as choosing informal regulation instead of government regulation. However, the contractual model most often chosen may, as mentioned, more correctly be described as a quasi-private ordering, as the contracts in many cases can be enforced by the legal system (Lemley, 1998), but the partners often choose not to do so.

3.3 When *de jure* rights are vague, what does this imply for data access?

General economic performance largely depends on the way property rights are defined. The ownership rights—of AI data here—therefore affect "three elements: (a) the right to use [data]; (b) the right to appropriate returns [from data]; (c) the right to change [data] form and substance" (Furubotn and Pejovich, 1974, p. 4). To this must be added the right to transfer these three elements to another person or organisation (Rubin and Klumpp, 2012, p. 205).

As discussed above (Sections 3.1 and 3.2), this type of data—machine-generated AI data—seems to fall between the cracks of the various laws. Legal ownership—an exclusive right—of machine-generated data simply does not exist. What does that imply for data access? Moreover, what are the alternative solutions for governing AI data access?

A consequence of the lack of clearly defined property rights is that price cannot be relied on as a sufficient allocation mechanism. Property rights are the very foundation for the price mechanism to work. With that, the price works as an invisible hand in the market, and there is a clear basis for production, exchange and distribution. The price guides the allocation of goods and resources to the highest valuing users. Firms and individuals can have confidence to invest in and to interact on a market guided by contract law. For this to work, the contracts regulating transactions must be, as Macneil (1974, p. 738) stated, "sharp in by clear agreement; sharp out by clear performance". The identity of transaction partners does not matter, and the agreements made from the outset are very clear. Contract law is then sufficient to solve any conflict (Macneil, 1974; Williamson, 1981).

A market—in a sense of the classic contract concept—therefore does not exist here, as property rights are not assigned. Then identity matters; relationship matters; *de facto* control matters. The upshot of the review of what is written about the formal rules of the institutional framework—in Section 3.1—is that in essence, there are several ways to obtain control, and we discuss them in the following paragraphs.

Burstein (2012), in discussing the difficulties of contracting over information—derived from the famous information (disclosure) paradox put forward by Arrow (1962)—has explored three solutions—besides IPRs—to promote a robust market for information. As we see it, much of Burstein's (2012) description of the characteristics of information is very similar to that of AI data. For example, information is "not always non-excludable", it "[occupies] a spectrum", it is "not always a homogenous asset", it is "complex and multifaceted", and it is "subject to some inherent limitations but also manipulable by its holders" (pp. 26-39). The solutions Burstein proposed for a viable market for information—contractual relations, norms-based mechanisms, and other business strategies—are therefore, at least partly, applicable to AI data. They can function as important remedies in AI exchange relations, alone or in combination, when *de jure* rights are vague (or lacking), and/or when IPRs cannot play a significant role in the transaction. In the following, we discuss four remedies, and the first three are in line with Burstein's thinking on market for information.

The first remedy is bilateral contractual agreements:

For an analysis of governance relations and strategies in sectors using AI and ML, it is not enough just to look at legal property rights as IPRs. There are both *de jure* rights and *de facto* control, and the distinctions between them are often not made in general microeconomics analyses.

De facto control, including control of use and appropriate returns, are often of an *in personam* character, which implies that identity is what matters. With *de facto* control, the access (to data) often involves a bilateral contract between seller and buyer. The market does not exist in a strict classical sense. Buyers will not have the freedom to resell as they do with *in rem* rights. Consequently, a reallocation to the highest valuing user will be difficult.

Here, when the *sui generis* protection (of AI data) is vague (as discussed in Section 3.1), data transactions are governed by *de facto* control of data, as much due to technical circumstances as described in Section 2. As a consequence, a bilateral contract is the favoured mode. The parties generally cannot strike a one-time bargain for the sale of AI data. When the output of the processed AI data is a form of knowledge that is difficult to patent or to protect by copyrights, long-term relationships are needed. Bilateral contracts are the means of transfer of accessibility with control.

A bilateral contract is a governance mode quite different from classic market transactions. The latter is based on established legal IPRs. An IPR like a patent solves problem with the information disclosure dilemma and gives incentives for innovation in a quite different way. The innovation can then be marketed and be an object of trade, so that it ends up with a higher valuing user. As in other cases of new knowledge production, there are often other actors who are better at marketing and producing. A patent at least lets the market process steer the allocation in that direction.

Machine-generated AI data, however, are not patentable/copyrightable in the present legislative framework (see also Dornis, 2021). To find a higher valuing user by a bilateral contract procedure seems to be difficult. It must involve a search process that is more cumbersome than that in a market-like process (where the market functions as an invisible hand).

The case of a likely development, or even a dominance, of a bilateral dependency relationship between supplier and user in AI data deserves special attention. In such a case, a long-term relationship is important. If the identity is to be preserved (i.e., vertical integration is not chosen) there must be checks of the type that characterise private ordering. Credible commitments that allow for adaptability should be visible in a long-term contractual relationship. A hybrid between market and firm will be chosen; there, the classic legal contractual conditions are not perfunctorily followed. Instead, informal institutions—such as trust and reputation—become important.

The lack of *de jure* rights to AI data has implications for liability. A lack of *de jure* rights has negative implications for clarity of ownership.[13] If ownership is confused, liability will also be confused. Wiebe (2017) illustrates the complications in an example of a self-driving car. All the data related to driving, driver, traffic and environmental are of interest. But who has the claim to the data? Is it the owner of the car, the manufacturer of the car, the state, the insurance company etc.? How can conflicting claims be handled? These are questions that must be addressed by the institutional framework if the data are to be efficiently allocated to different interested parties. The governance of data between different uses depends on how well the institutional framework stands up to the challenges of the new era.

When there is a lack of IPRs that allow the function of the market, the governance mode chosen can develop into a contractual relationship of mutual sharing of data. This coordination mode allows instant access to the data of other players. Drexl (2017) refers to a hearing of representatives from the industry, which showed little interest in new IPRs. Data sharing by means of contract was considered sufficient. Consequently, governance of rent allocations is not a market type. Rather the hybrid type—such as private ordering—is the preferred mode. This way of coordination applies to raw data (i.e., data at the syntactic level, like 0 and 1). For more processed data resulting in information and knowledge, there are reasons for relying on more marketlike type of governance mode, for the reasons outlined above.

The second remedy is norms and other informal institutions.

Studies of incentive mechanisms have long shown that in the absence of a clearly defined intellectual property, norms can support and regulate the exchange of information. This strand of literature includes Yochai Benkler's (2006) well-known book, *The Wealth of Networks*, which argues for "the commons" concept, namely peer production. Others, such as Fauchart and von Hippel (2008), prefer open innovation, arguing that there are other norms, rather than proprietary norms, that can support the exchange of intellectual products. They include norms such as reciprocity, learning and even reputation (mutual respect) that favour free exchange of information: idea sharing in a community is of primary importance.

Typically, the mechanisms that underlie these intellectual assets production system are based neither in markets nor in hierarchies (Burstein, 2012).

The third remedy is other appropriability strategies including control of downstream complementary assets.

[13] See note 5 above,

In the management literature, there are discussions related to the situation that property rights are not clearly defined (e.g., subject to continuous or occasional reassignment), and/or not easily to be implemented (e.g., in some developing countries). Then there is a risk for innovators to be in a "weak appropriability regime", as Teece (1986) formulated. When strategic concerns come into the picture, firms tend to search for controls for downstream complementary assets (in, e.g., manufacturing, marketing, services). There are hot debates not only about whether AI data qualify for copyright protection, but also about whether data used for training machines can be categorised as fair learning in the United States (cf. Lemley and Casey, 2020). So, there are clearly appropriability hazards here relating to this legal uncertainty. Consequently, controls on complementary assets are needed as the profiting from innovation framework (Teece, 1986) suggests. The types and forms of complementary assets—in the AI data access and (re)use context—however, remain to be explored.

The fourth remedy is more hierarchical governance solutions: the received wisdom is that firms tend to adopt more hierarchical governance modes when (legal) protection is weak (Oxley, 1999). The endpoint is hierarchy, where the invisible hand—the market—is supplanted by a hierarchical relationship (vertical integration).

4. Governance structure: a transaction cost perspective

Given the institutional framework mentioned above (Section 3), what does the AI data coordination look like? In the following we discuss (factors affecting) the choice of governance structure in an AI context. The same premise as the transaction cost theory applies, namely, we assume that economic activity is assigned to the firm, or to the market, "to economize on transaction costs" (Williamson, 1975, p. 21). That is, the governance structures are assumed to economise on the transaction costs.

4.1 The two behaviour assumptions

In the original foundations of transaction cost theory (Williamson, 1975), two human factors (bounded rationality and opportunism) and two environmental factors (uncertainty and small numbers of exchangers) are discussed, as they relate to exchange relations. How does this work in an AI context?

Bounded rationality means that the capacity of human beings to formulate and solve complex problems is limited. The pairing of bounded rationality with uncertainty/complexity is one reason why arms-length contracting can be costly. The limit of human actors in bounded rationality needs to be updated to deal with the joining of new agents—intelligent machines—that have the capacity to make decisions in a humanlike fashion but are not steered by emotions, as humans are. Nowadays, chess machines beat even the cleverest chess players. In other words, we now have machines with AI that can act as new agents. Or perhaps a more modest formulation is so-called collective intelligence: people and computers think together as super minds (Malone, 2018).

Opportunism means that human beings cannot always be relied upon to reveal all information pertinent to a transaction candidly. From time to time, people will try to take advantage of information asymmetries to exploit situations to their own advantage at the expense of the interests of the other contracting parties. Hence, opportunism is viable if there is asymmetric information. The question posed here is again to what extent is asymmetric information present in a world with AI? What happens, for example, if both supplier and customer are equipped with AI? Are we reaching a different level of information asymmetries, or just a different form of information asymmetries, in an AI world? This is not the least because machines are much faster than human beings at processing information.

Opportunism—in Williamson's transaction cost theory—is bounded by human actors. Both agents and principals can be dishonest, for example by distorting data, or confusing transactions. In the AI era, Wagner (2020) argues that there is a general transition from *homo economicus* to *machina economica*, which changes the principal-agent structure.

Homo economicus (of classical theory), also called the pure economic man, refers to the observation that economic/social outcome is the result of interactions among utility-maximising individuals who make rational decisions (Buchanan, 1969[14]; Kirchgässner, 2014). *Machina economica* refers to the evolution that AI agents too are designed to be economic actors, and they are faced with bounded rationality (Parkes and Wellman, 2015).

As we see it, this transition—from *homo economicus* to *machina economica*—can reduce the problem of bounded rationality, of opportunism, but not necessarily of information asymmetry. AI increases connectivity between humans and changes the ways people interact (Turkle, 2017). "Nowadays there is a computer in the middle of virtually every transaction" (Varian, 2014, p. 27). Human and machine are inseparable. Collective intelligence emerges. People and computers think together, with superminds (cf., Malone, 2018). Such superminds lead to a new type of principal-agent problem, not the least due to changes in the scope, scale and structure of principal-agent relationships.

Wagner (2020) pictures the emergence of a new, a triangular principal-agent relationship (see Appendix 1). There are three (rather than two) involved actors—two humans and one robot/machine—named user, AI provider and AI agent, respectively. The human user (of AI) functions as a principal, while the AI agent functions as an agent. What is special here is the dual role of the AI provider: as an agent because it supplies AI services to the user and as a principal because it owns the AI agent.

On the one hand, the AI agent (e.g., Siri, Alexa) does not have to be constrained by bounded rationality to the same extent as human beings. On the other hand, information asymmetries still exist, and so do the diverging interest/incentives. They are the classic grounds for opportunist behaviour.

Arguably, information asymmetries have reached a new level. There is "an unprecedented scope and scale of information asymmetries" (Wagner, 2020, p. 120). AI machines work not only faster, but also in a directed manner, as words like dataveillance (van Dijck, 2014, 2018; Degli, 2014; Danaher, 2018) illustrate. Moreover, how AI agents arrive at decisions is no longer (easily) traceable, transparent or explicable (cf. Doshi-Velez et al., 2017). Arguably, while AI has reduced the search cost on some fronts (for a tailored convenience), it has also aggravated information asymmetries on other fronts, as terms like asymmetric collusion (Schüll, 2014) illustrate.

[14] "The pure economic man must behave so as to take more rather than less when confronted with simple monetary alternatives. He must maximize income-wealth and minimize outlays. He must maximize profits if he plays the role of entrepreneur" (Buchanan, 1969, p. 38).

Going back to Wagner's (2020) proposal of an AI triangular principal-agent relationship and linking it to opportunistic behaviour, there is still a problem. AI services can be provided at zero cost (no charge). The revenue can come from other sources, for example, advertising. By profiling the user, it is possible for the human *AI provider*—together with an *AI agent*—to influence the user's purchasing behaviour to make the platform more attractive for advertisers. This can be considered as opportunism in that it takes advantage of information asymmetries in a way that benefits the AI provider. This is a classical type of principal-agent problem, where the user is the principal and the AI provider is the agent. There are similar principal-agent relationships between the human AI provider and the AI agent (e.g., the platform). Given its access to a huge amount of data and superior capacity to learn from these data, the AI agent is likely to have more information than the provider. The AI agent can use its informational advantage in way that does not benefit the AI provider (principal).[15]

As to the two environmental factors—uncertainty and small numbers of dealers—discussed in Williamson (1975), the extent to which that uncertainty—in the form of, for example, unforeseeable events—will change in AI time is difficult to predict.

Uncertainty here, in the AI data access context, is associated with changes in (the scale, scope and structure of) information asymmetries as discussed above. This can affect the transaction cost in, for example, controls in incomplete contracts. As Hart (2017) argued, some aspects of the investment are not contractible, or they are costly to contract on (e.g., innovation investment). This change can aggravate *ex post* bargaining inefficiencies (known as post-contractual opportunism) and consequently increase transaction cost.

The small number (of dealers) problem also exists, particularly in platform cases. On the one hand, these two-sided matching platforms (Uber, Google) use AI and big data techniques to help firms to find better transaction partners (through, e.g., AI-driven recommendations). On the other hand, users become dependent on these new global intermediaries—platforms—to find transaction partners, giving rise to a small number bargaining problem (Barach et al., 2019). Terms like platform capitalism and algorithmic profiling speak to this problem. Marciano et al. (2020, p. 348) point out that one of the platforms' main features is "that of preventing the gathered private information from being publicly revealed in the open market".

The small number of dealers problem can be particularly relevant in the embryonic stage of AI, when data coordination is still dominated by bilateral contracts and technical means. By the embryonic stage, here we mean two interlinked conditions: (a) contracting between vertically connected actors hitherto dominates and (b) data processing is hitherto more concentrated in the first two stages than in the latter two, if we use the Edinburgh categorisation on AI data processes: data collection, data preparation (and linkage), data access, data retention and reuse (cf. Edinburgh Report, 2020). In smart cities, for example, large data are generated relating to energy consumption and mobility by smart grids, smart sensors and internet of things technologies. Much infrastructural data can be in the public domain, and bilateral contracts have hitherto been sufficient to solve private actor-involved contracting.

[15] In other words, an AI algorithm enables an information directionality (bias) that is aggravated by network effects via digital platforms (Steinmueller and Mansell, 2020).

The above four factors—two human and two environmental—can have profound effects on governance structures, which we discuss below.

4.2 AI asset specificity

The choice of an efficient governance structure (i.e., market-hybrid firm) also depends on asset specificity, which describes the extent to which the investments are specialised to a particular transaction. Dependency between AI data suppliers arises because of investment characteristics (asset specificity) as well as market power (caused by, e.g., network effects). Dependency due to scale, however, does not have to be bilateral.

Asset specificity differs from similar relationships in the sense that it refers to the case where firms are mutually dependent on each other because of the existence of assets that are more valuable if they continue to transact with each other than if they transact with other firms. The transaction-specific assets can be of a different nature. Assets can, for example, be transaction-specific due to knowledge that is especially useful in a transaction between a certain pair of firms, equipment like machinery that is important for the needs of a certain customer, being close to a customer etc.

The upshot is that the profits of mutually dependent firms depend on how important the continuation of a transaction relationship is and the agreed-upon transaction terms. The price charged for deliveries is an example of an important transaction term. A supplier that knows that a customer is completely dependent on its deliveries can, by charging a higher price, appropriate a larger portion of the joint profits. Charging a higher price can be opportunistically motivated by falsely referring to new circumstances (not covered by the original agreement) that the customer cannot check. It is exactly in such cases that hybrids (forms of governance) are chosen like private ordering, where there are checks and balances like credible commitments and trust that complement written contract terms. It is a hybrid, as the price as such is not a sufficient parameter in the choice between terminating or continuing the business relationship. There is too much at stake. There is a need of adaptation to changing circumstances at the same time as a vulnerability to opportunistic behaviour.

In the AI era, asset specificity due to knowledge no longer has to be of a human type. It can, as well as being a result of human intelligence, be a result of AI. The different types of asset specificity referred to by Williamson (1981, p. 555, 1996, p. 60) are of six types: site specificity (specific advantages of being located close to each other like being able to economise on inventory and transportation costs), physical asset specificity (such as specialised dies that are required to produce a component), human asset specificity (which arises from learning by doing), dedicated assets (discrete investment in a capacity just to serve a specific customer), brand name capital (for example, can a manufacturer's brand name capital be dependent on the quality of services it provides in the succeeding distribution stage) and temporal specificity (where timely delivery is important). The first three types were listed by Williamson as early examples of bilateral dependency. Of special interest here is human asset specificity, which arises in a learning-by-doing fashion. This is exactly how AI works. Consequently, we can add AI data specificity to the list.

Hence, with AI and ML, asset specificity comes in the shapes of:

a) Human asset specificity (learning by doing in choosing and designing the algorithm; sometimes manually sorting the data categories).

b) Physical asset specificity (fixed investments in millions of robots/sensors capturing the raw data and learning by doing in a training algorithm).
c) Temporary specificity (data changing all the time in 3Vs—volume, variety and velocity).
d) Networked data specificity (the mutual dependence of price, outcome etc. in, e.g., autonomous driving or the smart city and comparable cases).

Arguably, there is an *AI specificity*, which is a sum of these assets, as AI can acquire knowledge that is especially valuable in transactions with a specific supplier/customer, as the human mind can.

AI specificity here can be illustrated in *smart cities* and *autonomous driving* cases: assets can progressively become more specific and mutual dependency can become higher and higher in the sense that alternative sources of data supply do not exist at that time/location (demand changes dynamically, and so does the necessary data). Exchanges then take on a progressively stronger bilateral character. Classical market contracting does not work here, as assets (e.g., data) are so specific to the trading parties. This form of collaboration, which exists now, may hinder the long-term development of AI and ML, however.

4.3 Choice of governance mode in the AI era

How are AI data coordinated in terms of choice of governance mode?

Data access relies either on in-house supply of data or on transactions with external partners. Decisions on whether to define the outer boundaries of the divisions need to be made.

In the following, we assess factors affecting the choices of governance modes, of which (hierarchical) firms and markets are the two extreme alternatives, from the transaction cost perspective associated principally with the works of Coase (1936) and Williamson (1975, 1985, 1991, 1996). Central to the analysis is firms striving to choose a viable governance structure that helps to minimise transaction costs. As Williamson (1981, p. 548) put it, "economizing is accomplished by assigning transactions to governance structures in a discriminating way." How can the choice of governance structures—hierarchies, hybrids and the market—be explained by efforts to keep down transaction costs throughout stages the from raw material (AI data) input to final consumer product (also data here)?

The following discussions, in the AI data access context, are also based on two assumptions: (a) profit maximation, in the sense of keeping transaction costs down (i.e., governance structures with better transaction cost economising properties win) is good and (b) actors involved in coordination are constrained by existing technical and institutional frameworks that characterise AI and ML (discussed in Sections 2 and 3).

In the following, we move back to the issue of governance structure by examining transaction dimensions. The different choices of governance mode available for coordination of activities in different stages in a value chain begin with gathering raw materials and end with placing final (consumer) goods in the market (see Hayek, 1945 on the price system; Coase, 1937 and Williamson, 1979 on organisations/firms). For AI data, this implies a chain from raw AI data in the form of bits and bytes information to the processed-derived data (or apps) after segmentation, matching and targeting. This process involves various sources and (external) coordination partners.

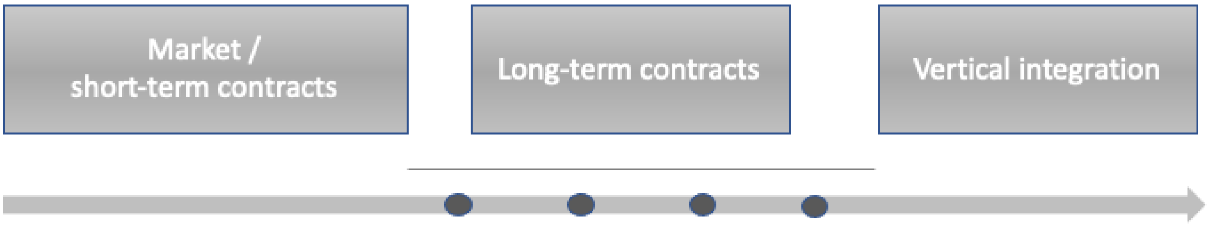
In the neoclassical market, price serves as the sole coordination mechanism, and continuation of a specific business relationship is not that important. There is no hesitation to use legal means to enforce written contract terms. This is a transaction situation that corresponds to the standard definitions of a market of perfect competition (pure competition) (see, e.g., Baumol, 1965, pp. 311-312; Scherer, 1970, p. 10). Transaction characteristics in this type of market include:

- Standardised commodities.
- Homogenous transactions (the identity of the trading partner is irrelevant).
- No entry and exit barriers.
- Many sellers and buyers and rivalry behaviour.

In such a case, the costs of the referred 3Cs—contact, contract and control—are kept to a minimum. Short-term classic contracts of spot character are sufficient. In the AI context, this is viable only when AI data are homogenous and when no contingent claims appear in the implementation.

Long-term contracts are a different kind, on a linear scale as Figure 1 indicates, and they can be put in a position between the market in the neoclassical sense and the firm. These are contracts in which the business relationship and asset specificities matter. Asset specificities refers to the investment characteristics that determine the degree of bilateral dependence (see Section 4.2). Bilateral dependence creates bottlenecks that can be strategically taken advantage of to increase profits at the expense of another trading party. Credible commitments, to the effect that both parties have something to lose if a business relationship is broken, are therefore often made to prevent strategic opportunism. These types of contractual relations go under the name private ordering (e.g., long-term contracts). Efforts are made to preserve the relationships and to find acceptable solutions for both parties rather than going to court in case of disputes. The goal is to make the contractual relationship adaptable to changing circumstances. Parties can rely on both informal and formal institutional constraints. Formal constraints as legal enforcement are a last resort if other means to preserve a relationship fail.

Figure 1 The market-hybrid-hierarchy continuum



Private ordering means that the participants do not rely completely on contract law. The pro is the flexibility in handling unpredictable situations. This is important in long-term relationships when there is a mutual dependency and credible commitments matter. The main con is that it cannot rely on law, as in other types of contractual relations, when solving conflicts. However, the credible commitments in long-term contractual relationships like private ordering might not be sufficient to solve conflicts "in a world where (at least some) parties are inclined [to be] opportunistic" (Williamson, 1979, p. 237). If these conflicts tend to be large, it might be better to have common ownership. In the AI data access context, this governance mode seems to be dominant, along with technical means (which introduces factual exclusivity).

Finally, vertical integration is when transactions are internalised through common ownership that makes it possible to supersede price as an allocation mechanism with authority (order). It is a governance mode in which interdependencies are so strong that the transaction cost of an arms-length contractual relationship like a market is too high. In Figure 1, the three different categories are put on a linear scale running from a purely competitive situation with price as the allocation mechanism to vertical integration, where the price can be superseded by (internal) authority in resource allocation.

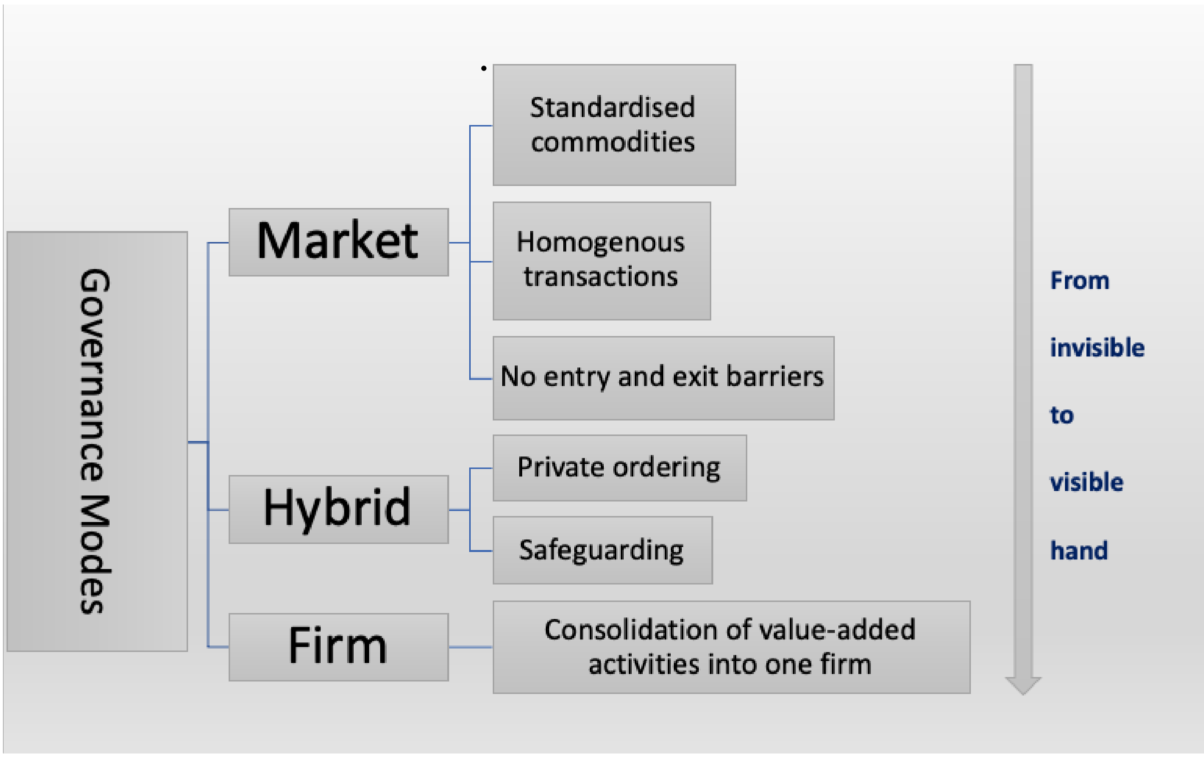
Transaction cost considerations are predicted to be behind the choices between these three categories of governance mechanisms. Of these alternatives, vertical integration is the most expensive in terms of coordination costs. The cheapest is the market alternative (the invisible hand alternative). A precondition for hybrid and vertical integration (hierarchy) is the existence of the bilateral dependence that can result from asset specificity. Uncertainty/complexity increases the contract cost of a hybrid solution.

As mentioned above, in pure competition, with all the assumptions stated above fulfilled, transaction costs are low. The quality of the goods and the identity and specifics of the transaction partners are well known, which lowers contact costs. In addition, the price is fixed, making both contract costs and control costs (enforcement and monitoring costs) low, as there are many attractive possible transaction partners.

As long as the supplier and the user are separate actors, there are conflicting profit incentives (i.e., the supplier wants as high price as possible, and the user wants as low a price as possible). This is no problem in a pure competitive market, as the price is fixed. However, as soon there is a bilateral dependency, there can be a scope for price negotiations after a contract has been made, and that can lead to high transaction costs. In a bilateral trading relationship due to, e.g., asset specificity, there are transaction costs, different types of contact, and contract and control costs to be considered when crafting a long-term contractual relationship. Contracts (pre-contract costs) are related to the investment decisions of different kinds that make the assets of firms interdependent. On one side, it is a question of the advantages of specialisation, making assets more productive if they are tailored to the needs of a transaction partner. On the other side, transaction costs also have to be factored in. There are implications for both contract and control costs. To avoid high control costs (post-contract costs) a contract has to give little scope for post-contract opportunism. Trust, norms, long-term relationships, and safeguards of different kinds that make it unprofitable to behave opportunistically in the future have to be considered. This is worth consideration, as long-term contracts by nature tend to be incomplete (Williamson, 1985, 1996). Future events are impossible to foresee perfectly, due to both complexity and uncertainty. This means that there is a need for adaptability to changing circumstances to consider in the contractual relationship. How trading relationships with specialised assets of human and/or physical assets involved in bilateral relationship of this kind are to be worked out is an important empirical research problem that has not received much study in the new era of ML and AI. Can separate ownership be what characterises governance? Or is it more of vertical integration with fewer conflicting profit incentives, but more internal control costs and problems of promoting incentives for employee productivity?

To sum up, Figure 2 shows the different factors that lead to transaction costs favouring one of the different governance modes (market, hybrid and firm).

Figure 2 Governance modes—from market to hierarchy



Source: Authors' own summary, based on Williamson (1975, 1981, 1985)

This has implications for AI data access. On the one hand, AI data are not homogenous and AI data creation is a complex process, involving AI data collection, data processing and application(s), with which investments in technologies, competences and infrastructures are made and multiple actors are coordinated. On the other hand, no features (of the data) or applications are *ex ante* given or known, implying dynamic changes in important transaction attributes—frequency, uncertainty and asset specificity—and consequently dynamic changes in governance modes. Looking at these three attributes, frequency matters, as AI is all about frequency. It is from frequency that the algorithms in a learning-by-doing fashion become more and more efficient and clever. Short-term contracts are not sufficient. There is a time dimension in access to data that implies long-term contractual relations. For every long-term relationship, there is uncertainty involved that makes contracts incomplete (Williamson, 1985). Not every future contingency can be foreseen. This can also invite opportunistic behaviour. However, when a bilateral dependency is created by asset specificity, the uncertainty inherited in long-term contracts invites parties to take advantage opportunistically at contract partners' expense.

To sum up, hypothetically, simple (crude) data can be exchanged in a market system (with, e.g., unit price); data can be contracted in a form of private ordering in situations when the coordination parties are *ex ante* known (e.g., smart cities). Data can be accessed in (hierarchical) firms with a high degree of specificity that in principle creates a strong mutual dependence.

5 Firm level appropriability

Firm-level factors such as resources, capabilities and strategies have important effects on how AI data coordination mechanisms behave. We make a distinction between appropriability and learning, namely between profiting from AI data (primarily in the private domain) and learning from AI data (largely in the public domain).

The premise adopted here is that firms act as nexuses of contracts, represented by lively decision-making individuals—varying in capabilities, resources, and strategies—who make the choice on different appropriability methods. One way to secure profit can be to get control over complementary assets (Teece, 1986).

It also has to be kept in mind that firms get involved in the AI data trade for multiple reasons. The industrial organisation alliances literature gives many explanations for this. For example, Kogut (1988) summarises three major explanations for why joint ventures exist: strategic, learning, and transaction cost. These mechanisms also apply to AI data coordination.

Appropriability in a digital context

For innovators, if IPRs are weak, a large portion of the innovation profits might spill over to others. In addition, if first mover advantage does not exist, innovations will result in dependency on specific suppliers/customers, which makes it important to choose a strategy focusing on complementary capabilities to prevent such spill-overs. Complementary capabilities including those in manufacturing, distribution and services are referred to as complementary assets by Teece (1986).

Weak property rights imply that contractual hazards are high. One problem is that protection of machine-produced data (knowledge) through IPRs is not effective (see Section 3). In court cases, the infringer of IPRs quite frequently wins (see, Cremers and Schliessler, 2015; Bjuggren, Domeij and Horn, 2015). First mover advantage can motivate entrepreneurs to invest in new knowledge even if IPRs are weak given that the (market) lead time over imitators could be long enough to enable a value capture (by the first mover). However, even if there is a strong first mover advantage, it can be difficult to appropriate entrepreneurial profits from innovation.[16] Specialised complementary knowledge and other complementary assets are important add-ons to first mover advantage, as they can ensure that profits are captured by innovators, rather than by other actors (Teece, 1986). These complementary assets can be in house or contractually sourced from external alliances. Because of the need for complements, various forms of alliances enter into the picture of strategic choice of appropriability means. This has also become a critical governance factor, as pointed out in several articles by Teece (1986, 2006, 2012, 2018). The strategic alliances literature—which initially described alliances in electronics, computers, software and telecommunications—has accordingly boomed since the mid-1980s (cf. Hagedoorn and Schakenraad, 1991).

Even with a strong IPRs regime, there is no guarantee that innovators will reap the main parts of the profits they create, according to Teece (1986, 2006, 2012, 2018). Instead, profits spill over to imitators and producers of critical components. These spill-overs are sometimes so large that the profits almost disappear in a short time. Teece's advice is to gain control over

[16] Not all first movers are based on innovations, but innovations are often an important source of first mover activity. For the simplicity, we treat first movers as innovators here.

crucial components by contract or ownership to safeguard profits. A new product or process innovation is likely to require complementary assets of different kinds that are tailored to the innovation. It can, for example, be new components produced by specialised machinery and equipment (physical asset specificity) or the specialised human knowledge required for the delivery of parts and marketing of output (human asset specificity). As long as these complementary assets are owned by other firms rather than by the innovator, a mutual dependency arises, and that must be handled to safeguard entrepreneurial profits. The degree of mutual dependency also varies due to the asset specificity. Transaction cost issues enter the scene. Alternatives are contracts of a hybrid type such as private ordering, partial ownership and vertical integration. Teece describes different arrangements of that sort made by innovating firms. Teece points out that creating a licence market for innovative knowledge is often not very profitable. To gain control over critical components through hybrids of means is often a better way to appropriate the profits of an innovation.

Teece (2018) stated that the movement of the digital economy continues along the same lines, namely that it is important for innovators to control complementary assets in the new platform era too. To prevent large parts of the profits from being diverted into other stages of the value chain, innovators have the downstream activities in house or controlled through partnerships. The types of partnerships – and transaction cost derived – are hinged with the types of investment. This was inspired by Williamson (1975, 1985), who emphasised the importance of handling transaction problems as they arise when there is a dependency due to asset specificity. The nature of assets in the new digital word may change. Some assets are firm-specific, like algorithms that learn faster and produce more output the more narrowly focused they are (so-called deep learning is promoted by a narrow focus). Special attention is paid to the character of a new sort of so-called general-purpose technologies that have developed in the new digital economy: platform technologies. Platforms are the focus. The definition of a platform is "any combination of hardware and software that provides standards, interfaces and rules that enable and allow providers of complements to add value and interact with each other and/or users" (Teece, 2018, p. 1375). Teece argues that complementary assets are becoming even more important in the new era of AI and ML. Examples of platforms in use are Alphabet and Apple, where vast amounts of data are collected that can be processed and used downstream. Hence, to create incentives for innovation, it is even more important to control and to appropriate the entrepreneurial profits, rather than letting them end up in the bottlenecks (between upstream and downstream) caused by a lack of indispensable complements. Business strategies for preventing that from happening are needed, through means like alliances or vertical integration (acquisition). One example is Apple, which has integrated its software iOS into its hardware in the case of the iPhone, i.e., chosen vertical integration as a protective means. Platform, in Teece's treatment, is an enabling context in his borrowing of the concept of general-purpose technologies (cf., Bresnahan and Trajtenberg, 1995), or a matchmaker, a two-sided market (Rochet and Tirole, 2005), through which firms are organising new types of upstream and downstream industrial activities.

Marciano et al. (2020), however, argued that in the digital platform era, upstream innovators—platform owners—play a somewhat different role than what is argued by Teece (2018). Marciano et al. argued that market transactions take place upstream from the platform. The producers of complementary products to the platform compete upstream. In other words, it is the platform that provides a competitive market. Examples are Microsoft and Apple, which provide markets for applications. The platforms themselves tend to make the firms that produce them dominant, benefiting from economies of scale and scope in supply, and from the network effect emerging from the demand side. The competition for supplying a platform is in this sense a competition for establishing a market. Upstream innovation here is not a single discrete product or service, but innovative ways (business models) of initiating new types (and scales and scope) of connectedness and interactions, namely a new market (e.g., Facebook). Platform, in Marciano et al.'s treatment, IS the upstream source in a collective form.

Furthermore, the role of first mover advantage is getting stronger attention in the platform era. From the appropriability perspective, the first mover advantage—a temporary Schumpeterian monopoly —applies here in platform case. This observation—on the importance of first mover advantage—is not new (see Levin et al., 1987; Cohen et al., 2000). Market lead, as a classic approvability means, is derived from the cost advantages in learning, which in turn are derived from the firm's stock of tangible and intangible assets (resources). There is always a dose of secrecy in market leaders. In fact, Penrose (1959) first identified the causal linkage between first mover advantage and firm' resources/capabilities. Lieberman and Montgomery (1998) argue that there is a reinforcing interplay between the timing of entry and a firm's knowledge stock, and Lieberman (2005) follows up with an empirical study of how the survival of firms in the 1990s dot-com chaos can be explained by first mover advantages.

The first mover advantage for firms is now, in the AI era, positioned in a network environment, and it applies to firms with patented innovation. One important part of the first mover advantage is network economies emancipated from the demand side (Goldfarb and Trefler, 2018). Connectedness is a key word in AI and ML (Porter and Heppelmann, 2014). Here, both economies of scale and of scope apply. (a) Scale: the more connected users/things there are, the larger the dataset that will be produced. Larger datasets will be assessed more accurately by ML algorithms (lower cost per unit), for example the price and availability of parking spaces in a smart city setting. Consequently, more users will want to join (network effect). (b) Scope: smart parking data can then be merged with data related to users' consumption patterns (e.g., willingness to pay at certain price levels). ML can then recommend reasonable parking trade-offs (price versus distance/time for further driving) to individuals. The larger and more varied the dataset, the more useful (multi-functional) the dataset will become to additional users, and the bigger the market this smart parking application will gain.

Consequently, a self-enforcement mechanism, a feedback loop, a cumulative causation mechanism might be in play. This means that this first-mover advantage may not ebb along an Utterbackian (Utterback and Abernathy,1975) product life cycle, or along an S-curve (Foster, 1986). On the contrary, there is potential for a prolonged natural monopoly for the platform owner, similar to what is dubbed the Hirshleifer case (cf. Hirshleifer, 1971) in which innovators can appropriate returns by timely moves to take a long position in innovation and in complementary assets.

The concept of asset specificity also needs to be addressed. Asset specificity has a role to play here, for example, in the choice of strategy to prevent large spill-overs from entrepreneurial firms. Complementary dependence upstream and/or downstream creates a situation of dependency that can easily result in large spill-over. To some extent, antitrust laws that prevent refusal to deliver—described in Section 3.1—can modify this picture too.

Moreover, the industrial utilisation of ML/AI technology is still in its infancy. While the balance between access and exclusion has been at the centre of the debate on the AI data regime on analytical levels, the focus now, in this infancy stage, is primarily on access. Empirical studies on industrial deployment of ML show that knowledge creation, technology and organisational coordination are interrelated: the value of data for training ML algorithms depends on access to others' data (to make the processed data representative of the real-world population). Many firms are still in the exploration stage out of developing data-based business models and sensible collaborative modes for data access and sharing, illustrated by the (pan-European) smart parking case and the airport speech recognition chatbots case (Long and Grafström, 2021).

To be in the game—the bandwagon effect of (co)creating AI data—is vital, particularly when product characteristics such as data quality and data coverage are concerned at this infancy stage. Data quality is related to the accuracy, completeness and interoperability of the data,[17] and data coverage includes both geographic and demographic coverage of the real-world population.

Arguably, the current EU legal instruments—the General Data Protection Regulation (GDPR) related to personal data and the DBD related to database producers at the forefront—leave a wide legal no-man's-land for firms in data access practices (Martens, 2018). Current data access practices cover a wide span of strategies, ranging from free access to monopolistic control on the one hand, and on the other hand, access that is heavily ruled by bilateral contracts and technical protection (Martens, 2018). Data traits like open or closed data, level of standardisation (interoperability) and security are important elements at this stage, influencing choices of modes of data access.

Appropriability and AI data peculiarities

It is still very unclear whether market forces or influential regulatory interventions such as the GDPR and DBD have hitherto driven access and data trade (e.g., pricing), pushing the market beyond voluntary exchanges. Stakeholders have very diverse kinds of interests, and often, exclusion of data causes more drawbacks than benefits, particularly for small actors.

Two additional elements affect data access: (a) data in the hands of the public or private domains and (b) data in the hands of small and medium-sized enterprises and large (platform) owners. The platform owner was discussed in Section 2, in the context of data characteristics. So, we mainly discuss the public and private domains of data.

Whether data are in the public domain or the private domain plays a role. Openness can be due to mandatory portability as in the EU's GDPR Article 20: the right to have the personal data transmitted directly from one controller to another, or on voluntary bases[18] such as actors' sense of community, of commons. It is neither in the market nor in the firm, but it has the potential to facilitate economies of scope in data aggregation.

[17] The raw data often come in a variety of structures and formats, causing problems in interoperability.
[18] This can be found in the voluntary FRAND data governance rules proposed by C-ITS Working Group 6 (European Commission, 2016), in which OEMs are obliged to give access to any party with a legitimate claim to use the data (cf. Mueller-Langer and Martens, 2018).

Data in the public domain means that a permit of (re)use is not needed (cf. Zurth, 2020, p. 18). The distinction between private- and public-domain data is important for training algorithms, as AI is like humans in that it obtains knowledge through learning by doing, just as a pupil gets knowledge trough training and reading schoolbooks. To do this, there must be training data available. This is especially important for new (small) firms.

There are two types of arguments here, and they differ in the degree of openness.

The first argument is that "autonomously generated non-personal data should fall into the public domain. [They] should be open data, excluded from protection by the Database Directive (DD), the Copyright Directive (CDSM) and the Trade Secrets Directive (TSD)" (Kop, 2020, p. 1).

The second argument is that data are in a kind of quasi-open public domain. Within this argument, there are two branches.

One branch advocates a form of fair use, a phrase/doctrine/clause in the copyright protection exception categories. Mark Lemley and Bryan Casey (2020) introduce a concept of fair learning in AI: a database for algorithm training can be treated as having a non-expressive purpose, which should be permitted for learning purpose use by others. The logic goes: "training sets are likely to contain millions of different works with thousands of different owners, there is no plausible option simply to license all of the underlying photographs, videos, audio files, or texts for the new use" (p. 111). The learning purpose is phrased as "The ML system wants photos of stop signs so it can learn to recognize stop signs, not because of the artistic choices you made in lighting or composing your photo" (p. 112). It is, however, difficult to draw the boundaries concerning when data is used for training in practice.

Another option is to establish a so-called data sharing alliance—which is quasi-open, open for alliance (consortium) members but closed for the market. This can create *learning houses* in possession of training datasets to which individual firms can get access via licensing agreements (Cortez et al., 2020)

Regardless of these arguments—whether data used for training purposes should be in the public or quasi-public domain—there is already extensive data in the public domain. A large proportion of the current B2B AI data is an outcome of the process of extracting public-domain data (e.g., city maps used in smart city-related applications[19]).

Much of the data in the public domain, however, are of poor quality (not updated, not detailed enough, in a proprietary format and consequently with low interoperability), as we discovered in our pilot study in a smart city project. This raises a question: if private actors add more data (value) to the existing public domain data, enabling improved use of original public data, would then a so-called copyleft principle apply? Copyleft is a concept used in the open-source software community, meaning that the derivative works should follow the same doctrine as the original work (i.e., keeping open access). Weibe (2020) discusses this for scientific infrastructural projects related to academic works and their related data. In the

[19] A map, in the public domain, is the departure point for nearly all smart city applications (smart government, smart transportation, smart mobility, smart energy, smart building, smart home, smart healthcare, smart education, smart security…).

private sector, it is unclear which principle applies. Using the copyleft principle—namely if used data were open, then the derivative works also need to be open—sets constraints on the appropriability strategies private actors can use. In a networked environment, public-domain data (e.g., maps) are often a basis for building up applications further.

Another issue is that actors (firms) intentionally make their (upstream) data open, constructing an appropriability strategy with a focus on the downstream co-specialised asset positions (in, e.g., manufacturing, sales, marketing). This is particularly useful in situations where openness is the norm in industrial practices, or in situations where the upstream data (research) are not subject to IPRs protection (not patentable). Two industries stick out here. One is the software industry, in which open-source codes are commonly used. Another one is the biomedical (drugs) industry, in which a strong downstream asset position in the sale and marketing (of drugs) is important, because the upstream discovery (on, e.g., genes associated with a particular disease) is often not patentable. Pissano (2006) discusses both cases in his revisit of Teece's (1986) profiting from innovation framework.

In the open source-based software industry, for example, much of the software used for data processing is open source (e.g., ojAlgo[20]), and hardware can be purchased via cloud services. Openness is therefore an embedded element, and it varies only in degree in many industry sectors. Pisano (2006) argues that the advantage is of a dynamic character where, for example, a source code is made open freely without any cost if the attached norms are followed. In the platform era, there are two advantages: firstly, there is feedback from users that helps to develop the code or other types of new goods or processes (Pearce, 2017). Therefore, the cost of research and development can be reduced, and the feedback can lead to a comparative advantage for the originally innovative firm. Secondly, firms can choose to open their works (e.g., software code) strategically to enable complementary service provision around innovations. This is particularly viable in a weak appropriability regime in which openness is the norm (Pisano, 2006). Pisano and others cite the example of Linux, which had repercussions for important digital actors like Microsoft's and IBM's later choices of open access software. Another related aspect mentioned by Hartman and Henkel (2021) is that human assets are an especially important resource in the new era. Recruitment of top scientists and highly educated academics is a priority. For top scientists, publications are important, and they want their publications to be read and cited. Open access can therefore help in the recruitment process.

Data in the private domain—which is the focus of this paper—have hitherto been accessed primarily via two means: (a) technical protection means (e.g., authentication) or (b) bilateral contracts (e.g., data-sharing agreements). Both methods introduce *factual exclusivity* (Martens, 2018), with problems in scaling up the coordination needed for training algorithms.

The former—technical protection means—may be considered a kind of unilateral contract (e.g., a simple formed license), which hinders a further transferability of rights, for example in re-use of AI data. The latter—bilateral contracting—may be perceived as a kind of relational contracting, a hybrid governance mode in between market and hierarchy in Williamson's (1979, 1985) terms.

Typical arguments here include relational contracting being suitable when uncertainty is high (Crocker and Masten, 1991). Uncertainty here also embraces the legal uncertainty mentioned

[20] https://www.ojalgo.org.

in Section 3. This can help actors to handle unanticipated contingencies with which the classic (market-based) contracting cannot deal (Williamson, 1979). This also applies in (relational) contracting practices including enforcing behaviour norms between transactors, namely norms such as information sharing routines and long-term reliance play important roles (Macneil, 1980; Goldberg and Erickson, 1987). The *ex-ante* contracting costs may be lower, but actors may incur significant *ex post* bargaining costs as they periodically negotiate contract adjustments.

6 Summary

AI and the internet of things represent an industrial revolution. It is characterised, to a certain degree, by a replacement of humans by machines (algorithms) that handle huge amounts of data beyond humans' brain capacity. These algorithms are capable of learning like humans. As in humans, information is needed for training. Access to information/data enables machines to become increasingly clever. This paper is an attempt to decompose factors influencing how economic organisations solve the problem of data access and coordination in the new economy of AI and ML. The primary focus is on industrial data (B2B data) that cover the whole industrial process from data creation to the final consumer product. Our approach to the topic is inspired by new institutional economics as described by Williamson (2000). Seeing firms being constrained by its institutional framework, seeing firms as nexuses of contract, and seeing firms with individual variance in controlling their own resources, capabilities and strategies, we see AI data coordination logic analyses as structured along those elements.

The technical specificities of the new economy are described in this paper. The primary production input, data, is heterogenous, which makes organisation of transactions challenging. Data as a concept covers both unstructured and structured information. At the same time, data are dynamically changing, in volume, variety, velocity and veracity, and the value of one's data depends on one's access to others' data. Moreover, data are heterogenous, meaning that they can be mere information, value-added knowledge or property (commodity). Furthermore, there are scale advantages in both the supply and the demand (of data) to consider, as well as economies of scope. This causes problems in allocating access, control and rents in a networked environment.

Typical legal problems—as a consequence of the AI data characteristics described above—include how to determine the ownership of the data. To start with formal institutions, a division can be made between *in rem* and *in personam* rights, where *in rem* characterises property rights and *in personam* characterises contracts. Specific to AI, property rights are essentially lacking for data except for the use of trademarks and trade secrets. There is a big debate on whether the introduction of new property rights for data is necessary. On the competition (law) front, the scale and scope advantage of data provide advantages for first movers, often large platform companies. Remedies on the anti-trust front—notably the essential facility doctrine—are accused of joining in too late when a monopoly situation already exists. To sum up, this type of data falls between the cracks of the various laws, and there is a clear legal uncertainty.

So far, contract is the primary the legal means of protecting data. The lack of clearly defined (exclusive) property rights for data leads to confusion over ownership with consequences for liability among and rents in economic organisation. Norms and trust become important in transactions, making private ordering and hierarchy attractive modes of transaction. *De facto* control—through, for example, technical means—is also used complementarily .

From an economic organisation perspective, the question is how transactions—along the value chains from data creation/capture to final consumer product—are to be organised. Transaction cost analysis suggests a choice among three categories of coordination mechanisms, namely markets, hybrids and vertical integration, from the invisible hand of the market to more visible hands with the firm as the endpoint. This is, however, built upon assumptions about human characteristics. The basic assumptions about humans are bounded rationality and opportunism. These assumptions will not explain transactions clearly in the new economy, at least not to the same extent. AI is, arguably, less restricted by bounded rationality and less concerned about opportunism (than human agents are). It remains to be seen, however, how the programming of algorithms deals with that element.

According to TC, bilateral dependency is an important factor in the choice between markets, hybrids and firms. Bilateral dependence increases transaction cost in markets. The factor that gives rise to bilateral dependency is asset specificity. In AI, new types of asset specificity arise. Hybrids and firms become more attractive. The hybrid, private ordering, is described in several articles as a governance mode. Another important factor for the market to function like an invisible hand is that well-defined and enforced legal property rights exist. The institutional analysis shows that this is not the case in the new economy. So, what remains are hybrid and firms.

For innovative firms, like those in the new economy, it is important to safeguard profits from innovation. This is a reason for IPRs is to provide incentives for innovations. However, as Teece showed in several articles, IPRs are not as effective as physical property rights to protect owners. Imitations can be hard to prevent. Furthermore, dependence on the complementary assets of suppliers and customers often means that the profits of innovations end up with other firms. This is a consequence of the bilateral dependency described in transaction cost analysis. It is also another reason why using the market as a governance mode is avoided in transactions.

However, private production and access to data based on exclusivity and profit incentives is not all that matters in the new economy. An infrastructure of good character in the public domain is also important. It is important for training algorithms, just as school education is important for training children for their future lives. When it comes to non-personal data, there is no personal integrity to guard. Increased capabilities for algorithm learning can be achieved by voluntary cooperation between firms (data-sharing alliances) and by government investment on such things as up-to-date city maps. Firms are also deliberately making data and algorithms open source because of the advantages offered by such an institutional choice. One associated advantage can be that that it facilitates the recruitment of top scientists.

While the causal arrows can go both ways in appropriability and governance modes, the choice of governance modes is essentially an outcome of product characteristics, institutional setting and transaction (and appropriability) hazards in AI.

References

Agrawal, A.; Gans J.; Goldfarb, A (2019). Economic Policy for Artificial Intelligence, Chapter in NBER book *Innovation Policy and the Economy*, Lerner, J.& Stern, S. (eds.) University of Chicago Press. pp. 139 - 159.

Aghion, P., & Howitt, P. (1998). *Endogenous growth theory*. Cambridge, MA: MIT Press.

Alchian, A; Demsetz, H (1972). Production, Information Costs, and Economic Organization. The American Economic Review , 62 (5),777-795

Arrow, K.J. 1962, Economic Welfare and the Allocation of Resources of Invention, in NBER (ed.) *The Rate and Direction of Inventive Activity: Economic and Social Factors*. Princeton, NJ: Princeton University Press, pp. 609-25.

Arora, A; Fosfuri, A; Gambardella, A (2001). *Market for Technology: The Economics of Innovation and Corporate Strategy*, Massachusetts: The MIT Press

Barach, M; Kaul, A; Leung, M; Lu, S (2018) Small numbers bargaining in the age of big data: Evidence from a two-sided labor matching platform; https://www.researchgate.net/publication/329604208_Small_Numbers_Bargaining_in_the_Age_of_Big_Data_Evidence_From_a_Two-Sided_Labor_Matching_Platform

Baumol, W.J, (1965), *Economic Theory and Operations Analysis*, 2nd ed., Prentice Hall.

Bekkers, R; Verspagen, B; Smits, J (2002). Intellectual property rights and standardization: the case of GSM. *Telecommunications Policy*. 26 (3) 171-188

Benkler Yochai (2006) *The Wealth of Networks*. Yale University Press.

Bjuggren, P-O, Domeij, B. & Horn, A., (2015). Swedish Patent Litigation in Comparison to European, *Nordic Intellectual Property Review (NIR)*, no 6 , pp. 504-522.

Buchanan, J.M. (1969) Is Economics the Science of Choice?, in: E. STREISLER (ed.), Roads to Freedom: Essays in Honour of Friedrich A. von Hayek , Routledge and Keegan Paul, London, pp.47 – 64; reprinted in: J. M. BUCHANAN (1987b), pp. 35 – 50.

Burnstein, M (2012) Exchanging information without intellectual property. https://www.bc.edu/content/dam/files/schools/law/doc/patconpapers/Burstein%20Exchanging%20Information.pdf

Calabresi, G; Melamed, D (1972) Property Rules, Liability Rules, and Inalienability: One View of the Cathedral. *Harvard Law Review* 85 (6) 1089-1128

Coase, R.H., 1937. The Nature of the Firm, *Economica*, 4:386-405.

Coase, R. H. (1960), 'The Problem of Social Cost', *Journal of Law and Economics*, **3**(1): 1–44.

Colangelo, G; Maggiolino, M (2017). Data access and AI: Antitrust vs. Regulation https://ec.europa.eu/competition/information/digitisation_2018/contributions/giuseppe_colangelo_mariateresa_maggiolino.pdf

Colangelo, G; Maggiolino, M (2017). Big Data as a Misleading Facility. European Competition Journal 13(2) 249. http://antitrustdigest.net/giuseppe-colangelo-and-mariateresa-maggiolino-big-data-as-a-misleading-facility-2017-european-competition-journal-132-249/

Cooter, R; Ulen, T (2008) Law and Economics. Pearson/Addison Wesley

Cortez, E, K.; Cardenas, J.; Kop, M.; Repousis, O, G.; Upreti, P, N (2020) Transatlantic Antitrust and IPR DevelopmentsBimonthly NewsletterIssue No. 1/2020(March 19, 2020) https://www-cdn.law.stanford.edu/wp-content/uploads/2015/04/2020-1.pdf

Cremers, K & Schliessler, P., (2015). Patent Litigation Settlement in Germany: Why parties settle during trial, *European Journal of Law and Economics*, 40:185-208.

Crocker, K.J., Masten, S., 1991. Pretia ex machina? prices and process in long-term contracts. Journal of Law & Economics 34, 69–99.

Danaher J (2018) Toward an ethics of AI assistants: an initial framework. Philos Technol 31(4):629–653. https://doi.org/10.1007/s13347-018-0317-3

Davis, L.E. and North, D.C. (1971) Institutional Change and American Economic Growth. Cambridge University Press, Cambridge. http://dx.doi.org/10.1017/CBO9780511561078

Degli, E (2014) When big data meets dataveillance: the hidden side of analytics. Surveill Soc 12(2):209–225. https://doi.org/10.24908/ss.v12i2.5113

Doshi-Velez, Finale and Kortz, Mason and Budish, Ryan and Bavitz, Christopher and Gershman, Samuel J. and O'Brien, David and Scott, Kate and Shieber, Stuart and Waldo, Jim and Weinberger, David and Weller, Adrian and Wood, Alexandra, Accountability of AI Under the Law: The Role of Explanation (November 3, 2017). Berkman Center Research Publication, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3064761

Doshi-Velez F, Kortz M (2017) Accountability of AI under the law: the role of explanation. Hg. v. Berkman Klein Center Working Group on Explanation and the Law. arXiv preprint arXiv:1711.01134. https://doi.org/10.2139/ssrn.3064761. alterantively https://arxiv.org/pdf/1711.01134.pdf

Dornis, T.W., 2021. Of ‘Authorless Works’ and ‘Inventions without Inventor’ – The Muddy Waters of ‘AI Autonomy’ in Intellectual Property Doctrine

Dow, G.K., 1987. The function of authority in transaction cost economics. Journal of Economic Behavior and Organization 8, 13–38.

Drexl, J., 2017, Designing Competitive Markets for Industrial Data: Between Propertisation and Access, JIPITEC 257-92.

Duch-Brown, N; Martens, B; Mueller-Langer, F (2017).The economics of ownership, access and trade in digital data. European Commission, Joint Research Centre Working Paper 2017-01. https://ec.europa.eu/jrc/sites/default/files/jrc104756.pdf

Edinburgh Report (2020): The Role of Data in AI: Report for the Data Governance Working Group of the Global Partnership of AI, prepared by Digital Curation Centre, University of Edinburgh. https://gpai.ai/projects/data-governance/role-of-data-in-ai.pdf

European Commission (EC) (2017) Building the European Data Economy, Communication to the European Parliament and Council.

European Commission (EC) (2018) Towards a common European data space, Communication to the European Parliament and Council.

European Commission (EC) (2020) Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions: A European strategy for data https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066

Emmanuelle Fauchart & Eric von Hippel, *Norms-Based Intellectual Property Systems: The Case of French Chefs*, 19 ORG. SCI. 187 (2008).

Evans, David S. and Schmalensee, Richard (2016), The New Economics of Multi-Sided Platforms: A Guide to the Vocabulary (June 9, 2016). Available at SSRN: https://ssrn.com/abstract=2793021 or http://dx.doi.org/10.2139/ssrn.2793021

Emmanuelle Fauchart, Eric von Hippel, (2008) Norms-Based Intellectual Property Systems: The Case of French Chefs. Organization Science 19(2):187-201.

Fauchart, E.; von Hippel, E. (2008) Norms-Based Intellectual Property Systems: The Case of French Chefs. Organization ScienceVol. 19, No. 2

Federal Trade Commission (FTC) (2014), Data brokers: A call for transparency and accountability, FTC, Washington, DC. https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

Foster, R (1986) Working The S-Curve: Assessing Technological Threats. Research Management 29 (4) 17-20

Furubotn, E.G; Pejovich, S (1974). *The economics of property rights*. Cambridge, Mass., Ballinger Pub. Co.
Galanter, M (1981) Justice in many Rooms: Courts, Private Ordering, and Indigenous Law. The Journal of Legal Pluralism and Unofficial Law 13 (19) 1-47

Gandomi, A; Haider, M (2014). Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management* 35, 137–1

Gawer Annabelle (2014) Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy* 43 (2014) 1239–1249

Goold, Patrick Russell, Artificial Authors: Case Studies of Copyright in Works of Machine Learning (November 20, 2020). 67 Journal of the Copyright Society of the USA (2020), Available at SSRN: https://ssrn.com/abstract=3734574 or http://dx.doi.org/10.2139/ssrn.3734574

Goldfarb, A. & Trefler, D., 2018, Artificial Intelligence and International Trade, Ch 19 in *The Economics of Artificial Intelligence: An Agenda*, Agrawal,A., Gans, J & Goldfarb, A. (eds.), The University of Chicago Press, Chicago.

Goldberg, V.P., Erickson, J.R., 1987. Quantity and price adjustment in long-term contracts: a case study of petroleum coke 30, 369–398

Hart, Oliver (2017) Incomplete Contracts and Control *American Economic Review 2017, 107(7): 1731–1752 https://doi.org/10.1257/aer.107.7.1731*

Hayek, F., 1945, The Use of Knowledge in Society, *American Economic Review*, 35(4):519-30

Hagedoorn J; Schakenraad, J (1992)Leading Companies and Networks of Strategic Alliances in Information Technologies. Research Policy 21(2):163-190

Hirshleifer, J., 1971. The private and social value of information and the reward to inventive activity. American Economic Review 61, 561–574.

Hodgson, G (2015) Much of the 'economics of property rights' devalues property and legal rights. Journal of Institutional Economics (2015), 11: 4, 683–709

https://www.cambridge.org/core/journals/journal-of-institutional-economics/article/much-of-the-economics-of-property-rights-devalues-property-and-legal-rights/2F18F9AEF246C28137C8C3133D69D2AB

Hohfeld, W. N (1913) Some Fundamental Legal Conceptions as Applied in Judicial Reasoning.*The Yale Law Journal*. 23 (1) 16-59

Hudson, R (1899), When Rights in Personam Give Rise to Rights in Rem. *The Virginia Law Register* , May, 1899, 5 (1) 13-19. Published by: Virginia Law Review. URL: https://www.jstor.org/stable/1097900

Jones, C; Tonetti, C (2020) Nonrivalry and the Economics of Data. *American Economic Review* 110 (9): 2819–2858 https://doi.org/10.1257/aer.20191330

Kirchgässner, G (2014) The Role of Homo Oeconomicus in the Political Economy of James Buchanan. CREMA Working Paper, No. 2014-01. Center for Research in Economics, Management and the Arts (CREMA), Zürich. https://www.econstor.eu/bitstream/10419/214547/1/2014-01.pdf

Kop, Mauritz (2020). Machine Learning & EU Data Sharing Practices

Kogut, B (1988) “Joint ventures: theoretical and empirical perspectives”. *Strategic Management Journal,* **9,** 319-32.

Stanford -Vienna Transatlantic Technology Law Forum, Transatlantic Antitrust and IPR Developments, Stanford University, Issue No. 1/2020
https://www-cdn.law.stanford.edu/wp-content/uploads/2015/04/2020-1.pdf

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. Application Delivery Strategies by META Group Inc. Retrieved from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data- Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lemley, M.A. 1998, The Law and Economics of Internet Norms, *Chicago-Kent Law Review*, 73(4):1257-1294.

Lemley, M.; Casey, B (2020). Fair Learning. Available at SSRN:
https://ssrn.com/abstract=3528447 or http://dx.doi.org/10.2139/ssrn.3528447
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3528447&download=yes

Lieberman, M.B. 2007, Did First-Mover Advantage Survive the Dot-Com Crash? WP, http://www.anderson.ucla/faculty/marvin.lieberman/Lieberman_IFMA.pdf

Lieberman, M.B. & Montgomery, D.B. (1998) “First-Mover (Dis) Advantag: Retrospective and Link With the Resource-Based View” Strategic Management Journal, 19(12)

Li et al (2020)IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, VOL. 7, NO. 1, JANUARY-MARCH 2020
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9025277

Long, V; Grafström, J (2021) “What prevents machine learning from transforming industries?” In Long, V and Holmén, M (eds) *Industrial Transformation and Technological Change.* pp 125-140. Abingdon: Routledge.

Malone T (2018) Superminds. The surprising power of people and computers thinking together. One world Publications, London

Macneil, I.R., 1978. Contracts: adjustment of long-term contract relations under classical, neoclassical, and relational contract law. Northwestern University Law Review 72, 854–905.

Macaulay, Stewart. (1963). "Non-Contractual Relations in Business. A Preliminary Study." American Sociological Review. 28:55-67.

Macneil, I.R., 1980. The New Social Contract. Yale University Press, New Haven, CT.

Malone T (2018) Superminds. The surprising power of people and computers thinking together. One world Publications, London. Also see https://cci.mit.edu/superminds/

Mansell, R; Steinmueller, W. E. (2020). Advanced introduction to platform economics, Edward Elgar.

Martens, B., 2018, The Importance of Data Access Regimes for Artificial Intelligence and Machine Learning, JRC Digital Economy Working Paper, 2018-09, EU.

Martens, B & Mueller-Langer, F (2018) Access to digital car data and competition in aftersales services markets. DOI: 10.2139/ssrn.3262807. https://www.researchgate.net/publication/328282496_Access_to_digital_car_data_and_competition_in_aftersales_services_markets

Marciano, A., Nicita, A. & Romello, G.B, "Big data and big techs: understanding the value of information in platform capitalism" European, Journal of Law and Economics, 50:345-58.

Merges, R.P. & Nelson, R.R., 1990, On the Complex Economics of Patent Scope, Columbia Law Review, 90(4):839-916.

Monteverde, K., Teece, D.J., 1982. Supplier switching costs and vertical integration in the automobile industry. The Bell Journal of Economics 13, 206–213.

Mueller- Langer, F; Martens, B (2018). Access to digital car data and competition in aftersales services markets. October 2018. DOI: 10.2139/ssrn.3262807. SSRN Electronic Journal

Nelson, R.R. & Winter, S.G. 1982. *An Evolutionary Theory of Economic Change*, Cambridge, MA: Harvard University Press.

Nelson, R.R. 1995, Co–evolution of Industry Structure, Technology and Supporting Institutions, and the Making of Comparative Advantage, *International Journal of the Economics of Business,* 2 (2) 171-184.

North, D (1986) "The new institutional economics." Journal of Theoretical and Institutional Economics, 142: 230-237.

North, D., 1990, *Institutions, Institutional Change, and Economic Performance*, Cambridge: Cambridge University Press.

OECD (2015) Data-driven innovation https://www.oecd.org/digital/ieconomy/enhanced-data-access.htm

Oxley, J (1999) Institutional environment and the mechanisms of governance: the impact of intellectual property protection on the structure of inter-firm alliances. *Journal of Economic Behavior & Organization* 38; 283-309

Parkes D, Wellman M (2015) Economic reasoning and artificial intelligence. Science 349(6245):267– 272. https://doi.org/10.1126/science.aaa8403 . Or with longer version: https://dash.harvard.edu/bitstream/handle/1/25622971/_aiEcon.pdf;jsessionid=18869485404BE6F8E626B5E1301221E8?sequence=1

Partha Dasgupta & Joseph Stiglitz, In*dustrial Structure and the Nature of Innovative Activity*, 90 THE ECON. J. 266 (1980).

Pearce, JM 2017 Emerging Business Models for Open Source Hardware. Journal of Open Hardware, 1(1): 2, pp. 1–14, DOI: https://doi.org/10.5334/joh.4

Penrose, E.T., 1959, *Growth of the Firm*, Oxford: Basil Blackwell.

Pilaszy, I. and D. Tikk (2009), Recommending movies: even a few data are more valuable than metadata, Proceedings of the 2009 ACM Conference on Recommender Systems, 93–100.

Pisano, G (2006) Profiting from innovation and the intellectual property revolution.*Research Policy*, 2006, vol. 35, issue 8, 1122-1130

Porter, M.E. & Heppelman, J.E, (2014) ”How Smart Connected Products Are Transforming Competition", Harvard Business Review, November. https://hbr.org/2014/11/how-smart-connected-products-are-transforming-competition

Rysman, M. (2009). The economics of two-sided markets. *Journal of economic perspectives*, *23*(3), 125-43.

Rochet, J-C; Tirole, J (2005)Two-Sided Markets: A Progress Report. http://idei.fr/sites/default/files/medias/doc/wp/2005/2sided_markets.pdf

Rubin, P.; Klumpp, T. (2012) Property Rights and Capitalism, chap 7 in Mueller, D (eds.) The Oxford Handbook of Capitalism DOI: 10.1093/oxfordhb/9780195391176.013.0008

Scherer, 1970,F.M., *Industrial Market Structure and Economic Performance*, Rand Mc Nally

Macneil, Ian R. 1974. "The Many Futures of Contract." University of Southern California Law Review 67 (May): 691-816.

Stepanov, I., 2020, Introducing a Property Right over Data in the EU: The Data Producer´s Right – An Evaluation, *International Review of Law, Computers and Technology*, 34(1):65-86.

Schneier, Bruce. 2007. Why ‘Anonymous’ Data Sometimes Isn’t.” Wired. December 12, 2007. Accessed Dec 27, 2020 https://www.wired.com/2007/12/why-anonymous-data-sometimes-isnt/

Scotchmer, S (1991) Standing on the Shoulders of Giants:Cumulative Research andthe Patent Law. Journal of Economic Perspectives, 5 (1) 29-41

Solow, R. M., (1957). Technical Change and the Aggregate Production Function, *Review of Economics and Statistics*, 39:312–320.

Schüll, N.D. (2014) Addiction by design. Machine gambling in Las Vegas. Princeton University Press, Princeton. https://doi.org/10.1353/tech.2014.0015

Teece, D.J., 1986, Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy, *Research Policy*, 15(6):285-305.

Teece, D.J., 2006, Reflections on “Profiting from Innovations”, *Research Policy*, 35:1131-1146.

Teece, D.J., 2012, Next-Generation Competition: New Concepts for Understanding How Innovation Shapes Competition and Policy in the Digital Economy' *„Journal of Law, Economics and Policy*

Teece, D.J., 1988, Capturing Value from Technological Innovation: Integration, Strategic Partnering, and Licensing Decisions, Strategic Management, 18(3):46-61.

Teece, D.j., 2018, Profiting from innovation in the digital economy: Enabling technologies, standards, and licensing in the wireless world, *Research Policy*, 47:1367-1387.

Turkle S (2017) Alone together. Why we expect more from technology and less from each other. Revised trade paperback edition, 3rd edn. Basic Books, New York

Thompson, J. (1967). Organizations in Action. New York: McGrow-Hill.

Utterback, J.; Abernathy, W (1975) A dynamic model of process and product innovation. Omega. 3 (6), 639-656.

Varian H (2014) Beyond big data. Bus Econ 49(1):27–31

Varian H (2019) Artificial Intelligence, Economics, and Industrial Organization, chap 16 in Agrawal,A; Gans, J; Goldfarb, A (eds.) T*he Economics of Artificial Intelligence: An Agenda*. University of Chicago Press. 2019. P. 399-422

Van Dijck J (2014) Datafication, dataism and dataveillance: big data between scientific paradigm and ideology. Surveill Soc 12(2), 197–208. https://doi.org/10.24908/ss.v12i2.4776

Van Dijck, J; Poell, T; De Waal, M (2018) The platform society: public values in a connective world. Oxford: Oxford University Press.

Wagner, D, N. (2020) Economic patterns in a world with artificial intelligence. *Evolutionary and Institutional Economics Review* (2020) 17:111–131

Wernerfelt, B. (1984) “A Resource-Based View of the Firm” Strategic Management Journal, 5:171-80.

Wiebe, A., 2016, ‘Protection of Industrial Data’ – A New Property Right for the Digital Economy? GRUR Int 2016 (10): 877–884.

Wiebe, A., 2017, Protection of Industrial Data: A New Property Right for the Digital Economy, *Journal of Intellectual Property Law and Protection*, 12(1):62-71.

Williamson, O.E., 1975. Markets and Hierarchies. Free Press, New York

Williamson, Oliver E. (1981) The Economics of Organization: The Transaction Cost Approach. *American Journal of Sociology*, 87 (3) 548- 577

Williamson, O.E., (1985), *The Economics Institutions of Capitalism*, New York: Free Press.

Williamson, O.E., (1985). Markets and Hierarchies. Free Press, New York

Williamson, O.E., 1991, "Comparative Economic Organization: The Analysis of Discrete Structural Alternatives", Administrative Science Quarterly, 36 (2) 269-96.

Williamson, O.E., 1996, *The Mechanisms of Governance*, Oxford University Press, New York.

Williamson, O.E., 2009, Opening the Black Box of Firm and Market Organization: Antitrust, Ch 2 in Bjuggren, P-O & Mueller, D.C. (eds.), *The Modern Firm, Corporate Governance and Investment, Cheltenham: Edward Elgar Publishing Ltd*

Rysman Marc (2009)The Economics of Two-Sided Markets, *Journal of Economic Perspectives. 23 (3) 125–143*

Zech, Herbert. 2016. "Data as a Tradeable Commodity." In European Contract Law and the Digital Single Market, Implications of the Digital Revolution, edited by Alberto De Franceschi, 51–80. Cambridge: Intersentia.

Zech, Herbert. 2017 "Data as a Tradeable Commodity – Implications for Contract Law" (September 2017). Josef Drexl (ed.), Proceedings of the 18th EIPIN Congress: The New Data Economy between Data Ownership, Privacy and Safeguarding Competition, Edward Elgar Publishing, Forthcoming, Available at SSRN: https://ssrn.com/abstract=3063153

Zurth, P (2020) Artificial Creativity? A case against copyright protection for AI generated works. Oct 7, 2020. *UCLA J. L. & TECH.*

Appendix

Appendix 1. From Wagner (2020), Figure 1, p. 119

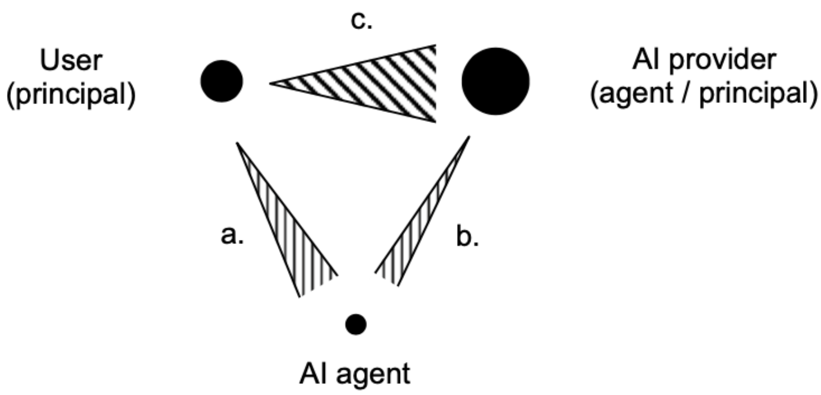


Fig. 1 Asymmetric information in triangular agency relationships with AI. The size of the circles symbolizes the type of actor: small = software, medium = human, large = organization. The angle of the triangles symbolized the degree of information asymmetry. Source: own representation