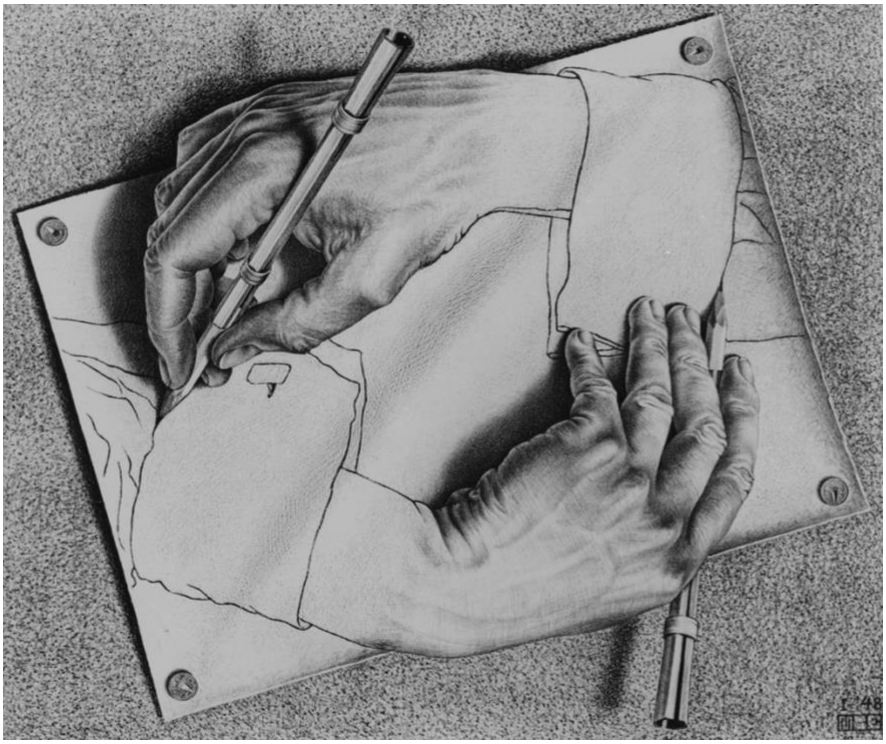
Master's Thesis – master Innovation Sciences

Intellectual property strategies in the age of artificial intelligence

An investigation of intellectual property strategies utilised by artificial intelligence start- and scaleups



Student
Name: Frank van Cappellen
Student number: 6622917
Contact: f.a.vancappellen@students.uu.nl

University supervisor
Name: Giovanna Capponi
Contact: g.capponi@uu.nl

Client supervisor
(NL AI Coalition & BG.legal)
Name: Jos van der Wijst
Contact: wijst@bg.legal

Course code: 2239X
Date: 13 August 2021

Abstract

Introduction: Artificial intelligence innovation, particularly by start- and scale-ups promises solutions to grand societal challenges and substantial financial returns. Despite substantially lowered barriers for entry to artificial intelligence development, uncertainties about intellectual property have inhibited innovative activity from accelerating. Past research has shown how well-designed intellectual property strategies tackle uncertainties, in turn amplifying innovation. Well-researched in related fields; software start- and scale-ups, little is known on intellectual property strategies utilised in the context of artificial intelligence. This abductive study is the first to investigate intellectual property strategies and the effects of innovation and market factors on its design in the context of artificial intelligence start- and scale-ups.

Theory: Extent research on software start- and scale-ups provided seven appropriability mechanisms: copyrights, database rights, trade secrecy, secrecy, lead-times, complementary assets, and technological complexity. Intellectual property strategies are affected by six factors, which differ between software and artificial intelligence. The type of key resources, vital to product development. Tacitness being the degree to which knowledge relies on skill and expertise Ambiguity being the observability of cause-and-effect relationships in knowledge. Open-source asset is the extent to which products rely on open-source assets. Market newness is the novelty of consumer needs and distribution channels to firms. R&D intensities is the size of research and development investments in relation to total investments.

Methods: The ways in which factors influence the usage of appropriability mechanisms utilised by artificial intelligence start- and scale-ups was investigated through cross-sectional qualitative data from nine semi-structured interviews, seven with chief executive officers, two with specialised artificial intelligence intellectual property advisors. Thematic analysis was employed to explore causal relationships between factors and appropriability mechanism usage.

Results: Findings confirmed industry similarities to validate the abductive approach and corroborated five theory-driven themes. R&D intensities provided contradictory results, confounded product maturity.

Discussion/Conclusion: The propensity for secrecy is concluded largest, albeit hampered by open-source motivations. Lead-times become increasingly relevant to keep up with rapidly changing consumer needs. As data and related assets increase in relevance, so does interest in database rights. However, artificial intelligence start- and scale-ups are deterred from its usage due to unclear legislative definitions. Insights provide a springboard for further research on intellectual property strategies in the context of artificial intelligence. Policy makers can improve database rights based on findings, better defining effort and the eligible of synthetic or derivate datasets. Uncertainties are addressed by examining their sources and illustrating the suitable strategies.

Acknowledgments

I thank my supervisor Giovanna Capponi and second reader Frank van Rijnsoever for valuable contributions. Also, I would like to express my gratitude to Jos van der Wijst, Anita Lieverdink, and the NL AI coalition for the assistance and access to data.

Contents

1. Introduction

1.1. Relevance

Artificial intelligence (henceforth AI) described by Marvin Minsky, co-founder of the AI laboratory at MIT, as *"the science of making machines do things that would require intelligence if done by men"* (Whitby, 1996, p. 1), is set to reshape industries and potentially change the world for the better. Greater adoption and integration of AI offers solutions to societal challenges such as climate change and energy poverty (Energy Central, 2020; TNO, 2020). Moreover, AI-assisted medicine is set to revolutionise cancer treatment (McKinney et al., 2020). AI is acknowledged by government, taking centre stage in the Dutch digital strategy proposal and is deemed vital for maintaining productivity levels in the face of aging populations (Rijksoverheid, 2020). It is then no surprise that in 2019 the AI industry was valued at 40 billion USD, with explosive growth rates of up to 42,4 per cent annually (Grand View Research, 2020).

AI is commonly mistaken as a recent technology, however initial conceptualisations such as symbolic learning, date back to 1854 with the introduction of Boolean algebra (Pamela, 2004). AI depends on prediction models, taught to recognise, and mimic patterns from large quantities of data, to produce predictions. Models for cancer recognition are provided images of healthy and unhealthy cells to learn patterns indicative of cancerous cells. Once taught, new images of cells are imported, letting the model make its own predictions. These types of models achieve accuracies upwards of 95 per cent, substantially higher than the 65 per cent average for radiologists (Svoboda, 2020).

Such practical use-cases have only recently started to emerge. This not for lack of human ingenuity, but rather limited computational capacity (Mucha & Seppälä, 2020). In the past acquiring sufficient capacity required large investments in specialised hardware, only viable for large well-resourced firms (Wagner, 2020). However, recent advancements in smaller and more affordable AI development hardware (Muehlhauser & Riber, 2014), development kits (Microsoft, 2021) and open-source repositions such as AutoGluon (Amazon, 2021) have lowered barriers for entry, resulting in a substantial influx of start- and scale-ups (henceforth SuSu's) in this sector (TechLeap, 2020). These relatively small, ill-resourced and young firms, are yet to find market fit or experience rapid growth in customer base (Heirman & Clarysse, 2004). Their small size allows for great flexibility, reorientating quickly in response to market developments (Miller Cole, 2019). However limited resources and business experiences leave little room for error, making these firms easy prey to large incumbents (Giardino et al., 2014; Paternoster et al., 2014).

1.2. Problem identification and gap in the literature

However, a study by Delponte (2018) for the International Data Corporation illustrated that, despite the substantial influx of SuSu's, AI SuSu innovative activity is yet to accelerate due to grave uncertainties surrounding AI-related intellectual property (henceforth IP). IP is defined as *"creations of the mind, such as inventions; literary and artistic works; design; and symbols, names and image used in commerce"* by the World Intellectual Property Organization (2021).

This study intends to address these uncertainties, in turn encouraging AI development by SuSu's. It focusses on AI SuSu's for several reasons. Firstly, past SuSu case studies in other fields such as hospitality (Guttentag, 2015), and mobility (Laurell & Sandström, 2016), have provided several novel theories and examples of innovation strategies. Similarly, novel insights might emerge by studying AI SuSu's. Secondly, AI is an emerging field with rapid developments in business practices. Focussing on front-runner taps into the cutting-edge of such industries (Ayoub & Payne, 2016). To this end, AI SuSu's serve as valuable sources of information, providing examples of the most up-to-date strategies. Third, AI development, especially by SuSu's, is expected to produce solutions for tackling societal challenges such as climate change, healthcare, and poverty reduction (Bradley et al., 2021), gaining a better understanding of suitable strategies has been shown to accelerate these processes (Teece, 2018a).

Most business cases rely on generating IP, in the form of proprietary assets, which are exploited for profit (Harrison & Sullivan, 2006). Uncertainties whether these assets are protected discourage investment stifling innovation (Teece, 2018a). These uncertainties are tackled by designing IP strategies (Teece, 2018a). Studies in law understand IP strategy as the process of aligning IP characteristics to existing formal mechanisms thereby simplifying litigation (Levine & Sichelman, 2018; Mattioli, 2014). Management studies define it as using routines and organisational structures in relation to external conditions to optimally capture value from IP (James et al., 2013a). Innovation and technologies studies define IP strategies as the "full toolkit of available mechanisms (and strategies) to capture value from innovation" (Pisano & Teece, 2007, p. 279). Firms design optimal IP strategies by understanding intellectual property regimes and the architecture of their industry (Pisano & Teece, 2007).

The seminal paper "Profiting from innovation" (PFI) by Teece (1986) demonstrated how market conditions influence IP strategies in the manufacturing industry. Other industries were later integrated, extending the PFI framework (Pisano & Teece, 2007; Teece, 2018b).

IP strategies are well-researched in traditional industries. An early case study explored causal relationships between market conditions and IP strategies utilised within 130 US business sectors (Levin et al., 1987). Later Cohen et al. (2000) gathered data on 1476 US R&D laboratories to further extend IP strategy knowledge. Lastly, Blind et al., (2006) combined these findings and applied them to data on 174 firms in eight of the largest industries in Germany to verify if findings found in America carried over to Europe.

A study on high-technology start-ups demonstrated how IP strategies in digital fields differ compared to traditional industries, implying industry conditions influence IP strategy design (Graham & Sichelman, 2016). Additional insights emerged from Miric et al. (2019) who studied developers of software SuSu's and linked differences in IP strategies to unique characteristics of digital fields. Findings from a study on digital platform developers corroborated these effects and specified challenges unique to digital fields such as ownership, ownership, value creation, and autonomies of complementors (Hein et al., 2020). Past research agreed that IP strategies differ substantially in digital fields, influenced by combinations of industrial, organisational, strategical, and technological characteristics.

In terms of AI, only a single study by Calvin et al., (2020) was found. The authors examined large US corporations and found these to be influenced by two simultaneous pressures. One, the pressure to patent for fear of counter suiting by competitors. Two, the pressure to open-source IP, to build a corporate image attractive to potential employees (who hold open-source motivations). Combined pressures led to a novel hybrid strategy, simultaneously patenting and open-sourcing IP (Calvin et al., 2020). While noteworthy, these findings are less suitable for the European SuSu's as these differ organisationally and legally. To this note, Calvin et al. (2020) suggest future scholars might focus on other AI stakeholders, to better understand IP strategies in the context of AI.

1.3. Research objectives and approach

This study has two objectives. First, explore which appropriability mechanisms are utilised by AI SuSu's. Second, determine the effects of factors on their utilisation.

"Which appropriability mechanisms are used by start- and scale-ups developing artificial intelligence products?"

To answer this research question pure inductive reasoning was considered ill-suited as the extensive body of work on IP strategies prevents one from approaching the phenomenon with no preconceived notions. Pure deduction was equally unfavourable as existing literature on AI SuSu's was preliminary and has yet to investigate IP strategies in the context of SuSu's. Instead, an abductive strategy is pursued, similar to Dubois & Gadde (2002) and Timmermans & Tavory (2012). This approach exploits existing concepts from a related field to investigate novel fields, while allowing novel concepts to emerge from data (Timmermans & Tavory, 2012).

This approach is best suited for situations in which *"the phenomenon is seen as similar to other phenomena already experienced and explained in other situations."* (Dubois & Gadde, 2002, p. 554). The field of software SuSu's was chosen following work by Kulkarni & Padmanabhan (2017) who found several similarities between software and AI SuSu's, particularly in the type of product, general product development, and the knowledge required for product development. Further validation came from Wan et al. (2020) who corroborated these findings and concluded AI SuSu's are "generally" similar to their software counterparts.

Numerous previous management studies validate its application for current purposes. For instance, Kindström et al., (2013) who studied the ways dynamic capabilities, originally based on product-centred firms, could be applied to product-and-service oriented firms. Or Høgevold & Svensson (2016) applied pre-existing theoretical frameworks on business sustainability efforts to novel data from Norwegian firms.

Thematic analysis similar to Fereday and Muir-Cochrane (2006) forms the basis for interpreting data. This type of analysis combines theory-driven themes, derived deductively from extent literature, with data-driven themes, generated inductively from raw data. First, based on extent literature on software SuSu IP strategies, a series of theory-driven themes were generated. These were combined to form a code manual, used for data collection and analysis. Next, the code manual guided nine semi-structured interviews with AI stakeholders. Interview data were coded using a code manual, assigning matching data entries to existing themes, and non-conforming data entries to inductively generated themes. Finally, themes were aggregated to form overarching findings.

Using software literature as a theoretical foundation contributes to science in four ways. First, extent software SuSu's theories that are shown to persist for AI SuSu's signify theoretical consistencies between the two streams of IP strategy literature, thereby bridging the literary gap. Second, extent theories that require alteration based on empirical data signify theoretical deviations between the two streams of IP strategy literature, further bridging the dearth of knowledge. Third, concepts emergent from data signal fertile points of departure for future research into AI SuSu's IP strategies. Fourth, applying existing theories to novel data validated extent literature, contributing to the greater body of SuSu IP strategy literature.

Also, answering the research question will result in several social benefits. Firstly, if successful findings will alleviate concerns voiced by AI SuSu's. Primarily by illustrating the sources uncertainties enables AI SuSu's to make more objective IP-strategy decisions, backed by science instead of emotion. But also, by explaining implicating factors for IP strategies lets AI SuSu's design more appropriate IP strategies, that account for any irregularities unique to AI. Lastly, issues shown to persist between the two industries can be tackled with existing management advice, directly offering AI SuSu's valid plans to improve their IP strategies. Moreover, findings can be used by policymakers to create suitable institutional environments meant to boost AI SuSu development. This has previously been linked to social benefits by inducing disruptions to locked-in systems, potentially revealing solutions to grand societal challenges (Bradley et al., 2021).

This study is conducted in collaboration with the Dutch AI coalition and BG.legal, providing access to data, funding, and valuable insights. The Dutch AI coalition or NL AI coalition is of comprised various AI stakeholders, including AI SuSu's, specialised law firms, and regional innovation hubs. Its purpose is to stimulate AI development in the Netherlands, which is achieved by hosting network events, giving workshops, and supporting research and pilot projects. Their main objectives include: building a strong knowledge and development ecosystem, tackling societal challenges, and creating economic opportunities for Dutch and European AI developers (NL AIC, 2021). BG.legal is an allied member and aims to support the AI ecosystem by building a knowledge platform titled: "Legal AI Rules & Regulations" (Legal AIR) (BG.legal, 2021).

The remainder of this study is structured as follows a section on theory, an overview of the empirical and analytical methodology, a presentation of results, final conclusions, and a discussion on the limitations and implications.

2. Theory

An initial framework, composed of a series of preconceptions is constructed from standing software SuSu literature. Preconceptions provided initial empirical and analytical groundwork for exploring the effects of various factors on IP strategies. Empirically they guided data collection towards concepts found relevant in similar contexts. Analytically they represented the lens through which data is analysed. Preconceptions are like hypotheses, predicting associations between dependent and independent variables. However, hypotheses remain static and are only tested during the final stages of research. Preconceptions on the other hand are dynamic, adapted based on novel discoveries emergent from data (Dubois & Gadde, 2002).

2.1. Appropriability mechanisms used by software SuSu's

Stefan & Bengtsson (2017) distinguished between three forms of appropriability mechanisms used by software SuSu's: formal, informal, and semi-formal. Formal mechanisms rely on legal documents to protect or appropriate profits from IP. Informal mechanisms rely on social structures, ensuring IP is protected and profits effectively appropriated. Lastly, semi-formal mechanisms rely on a mix of legal and social measures, ensuring returns on investments remain within the firm.

The most common formal mechanism used by software SuSu's are copyrights, used to protect source code, and at times databases (WTO, 1995). Acquiring these rights differs per region. In the Netherlands copyright is automatically assigned to original authors (de Laat, 2005). Whereas, US legislation requires registration before receiving copyrights (de Laat, 2005). Databases are usually considered illegible for copyrights because they do not represent conceptualisations of the mind. Instead, software SuSu's formally protect these assets via database rights (WTO, 1995). These rights are only attributed when substantial 'effort' was required to create the database (WTO, 1995). IP which is distinctly aesthetic such as user interfaces can be protected using design rights (Pike, 2001). Lastly, any aesthetic features not inherent to the product such as a logo or name, can be protected using trademarks (WTO, 1995).

Another example of formal protection mechanisms are patents. However, these see little usage by software SuSu's for two reasons. Firstly, the investments involved with attainting, maintaining, and enforcing patents are typically too large for ill-resourced SuSu's (Cockburn & MacGarvie, 2011). Secondly, European legislation does not consider software inventions patentable material, prohibiting its usage entirely (Hall & Ziedonis, 2001). Computer implemented inventions or CII's were introduced to solve this, requiring one or more components to be fulfilled via a computer program (EPO, 2021). However, required remain identical. For these reasons patents and CIIs are omitted in the current study.

Semi-formal measures, often in the form of non-disclosure agreements, are extensively used based on a study on 777 software start-ups by Levine & Sichelman (2018). These agreements typically describe information that must be kept confidential and lists the consequences involved when contracts are breached.

Alternatively, are several informal appropriability mechanisms. Most common is secrecy, encompassing all measures taken to ensure IP remains confidential (Cohen et al., 2000). Hemphill (2004) described how software SuSu's strategically manage secrecy by separating employees, ensuring no single person possesses all information (Hemphill, 2004). Additionally, policies meant to boost loyalty to ensure employees and their knowledge do not leave the firm are employed (Hemphill, 2004). Kitching & Blackburn (1998) found software SuSu's to comparatively prefer secrecy over other forms of protection. Anton & Yao (2002) explained this preference based on the unique characteristics of digital forms of IP. AI-related IP is of a similar type, intangible and difficult to capture in words. From this it is believed that AI SuSu's will make similar usage of secrecy.

At other times open-source development is utilised, defined as the production of software which is not reduced in utility by usage and from which access cannot be withheld by any organisation or person (Lerner & Tirole, 2016). Software SuSu's publish IP on publicly accessible depositories such as GitHub, allowing third-party developers to download and adapt original source code (Von Krogh et al., 2003). Publicly sharing IP does not imply open-source projects are entirely free to use as they are often protected by creative commons licenses (Creative Commons, 2021). These differ in two primary ways. Firstly, they differ in whether users must mention the name of the original author(s) (Creative Commons, 2021). Secondly, they differ in the degree to which users may commercialise products, developed using open-source software (Creative Commons, 2021).

There are several motivations for open-sourcing IP. Software SuSu's open-source IP to improve products by exploiting feedback provided by third-party developers (von Hippel, 2005). Other firms make products freely accessible to quickly attract large amounts of consumers to set an industry-standard, once products reach majority market shares (Von Hippel & Von Krogh, 2003). An example is fetchmail, which was codeveloped before becoming the standard for mail server hosting (Von Hippel & Von Krogh, 2003). Firms have also been shown to openly disclose IP to signal firm value to attract investors, clients, and potential employees (Calvin et al., 2020). At other times, IP is made open-source to pursue strategies involving complementary assets, developed either in-house of by third-parties (Hein et al., 2020). In-house complementary assets are typically used in "freemium" products (Pujol, 2012). These products come in two or more versions, a free version, with limited functionality and a paid version with complete functionality (Pujol, 2012). In this example, the latter represents complementary assets developed in-house. Users are first lured in with free access but will purchase premium versions, once limitations become too big of a nuisance, thereby generating returns from IP. Third-party complementary assets are typically used in "platform products", upon which third-party developers may freely build complementary assets. IP is appropriated by taking a percentage of revenues generated through these assets (Hein et al., 2020). Considering that AI is embedded in software and given the fact that open-source is deeply rooted in software development (Lerner & Tirole, 2016), it is likely that open-source strategies will carry over to AI SuSu's.

Lead-times, represents the competitive advantages provided to innovating firms as a result of relative higher rates of innovation (Cohen et al., 2000). Higher velocity of innovation dissuades competition from entering the market as they are unable to catch up with the innovating firm (Kessler & Chakrabarti, 1996). The resulting milder competition allows the innovating firm to generate additional returns from IP (Hilmola et al., 2003). Moreover, IP is protected from infringement since competitors are less inclined to steal continuously evolving IP (Hilmola et al., 2003). Software SuSu's generate these advantages by investing heavily in R&D to descend learning curves quicker compared to competitors (Hall et al., 2014). More specifically, software SuSu's have been shown to build lead-times by deploying updates in rapid succession, blocking imitators (Graham & Sichelman, 2016). Ample lead-time usage has been associated with rapidly changing or emerging software sub-fields (Paternoster et al., 2014). Based on this and suggestions by Kulkarni & Padmanabhan (2017), we expect lead-times to play a role in IP strategies in the context of AI.

Lastly technological complexity is used, whereby products are made unnecessarily complex to complicate reverse engineering (Samuelson & Scotchmer, 2002). Original products are dismantled by competitors to understand their innerworkings and built imitation products (Henry & Ruiz-Aliseda, 2012). Next, competitors sell their imitation products and by doing so generate returns on investments made by the original firm (Henry & Ruiz-Aliseda, 2012). Technological complexity in the context of software SuSu's involves compiling source code into machine code, consisting of 1s and 0s, difficult to decipher by humans (LaRoque, 2018). The consequence is that competitors cannot learn the innerworkings of original products, thus protecting against reverse engineering (LaRoque, 2018). Compiling is only usable for products reliant on source code. To this end, it would appear compiling becomes irrelevant in the context of AI because AI products do not rely heavily on source code. However, this is not to say that technological complexity becomes unimportant. In fact, Lauber-Rönsberg & Hetmank (2019) state that the unique characteristics of AI-related IP will likely result in new application of technological complexity. Thus, to capture novel applications technological complexity is included in the theoretical framework.

2.2. Software and AI: differences in IP definitions

An understanding of IP-assets in both contexts was required prior to investigating how these assets are protected. Definitions of software-related IP, written in Table 1, were formed based on the TRIPS agreement, the largest multi-lateral IP agreement active today.

Table 1

Forms of software-related IP

Definition
Software products;
Source code (full or partial);
Databases;
Distinct visually perceptible aesthetic features;
Distinct aesthetic features not contained in products (e.g., logo's, icons);
Undisclosed information receiving its commercial value by being secret.

Note. From WTO. (1995). *Overview: the TRIPS Agreement*. https://www.wto.org/english/tratop_e/trips_e/intel2_e.htm#relatedright

Determining AI-related IP definitions proved more challenging because AI has yet been included in TRIPS or comparable multi-lateral agreements. Instead, the standing body of AI-related IP literature was consulted. Literature focussed primarily on other subjects, such as highlighting in which areas AI is incompatible with legislation (Kop, 2020b). Or providing explanations as to why these incompatibilities occur (Hilty et al., 2020). Hence the standing literature could not provide clear definitions of AI-related IP.

Ultimately, AI-related IP was defined by first overviewing a standard AI SuSu business model and development process (provided below). Doing so illuminated which IP assets are most vital for effectively doing business in AI. Next, findings from this overview were applied to software-related IP definitions, written in Table 1. Variances between IP assets could be identified by directly comparing these lists, these possibly signalling factors affecting IP strategies.

AI developers begin by setting out product requirements including the required prediction and accuracy. Next, comes data collection, either real data; personal details, sensor data, or synthetic; procedurally created (Kulkarni & Padmanabham, 2017). Data processing follows where irrelevant entries are omitted and, in some cases, data is labelled. Data labelling involves manually assigning metadata, using experience and personal judgement to give meaning to data (e.g., adding tags to audio data, defining what dialect is spoken) (M Wachowicz & Gonçalves, 2019). Unlabelled data simply lacks such metadata. The decision for either relies on product requirements. Models based on labelled data will be able to predict any dialect assigned during labelling. Models based on unlabelled data generate pattern categories (gender, mood, etc.) and makes predictions based on these. Finally, data is split into training and testing sub-sets. Following, AI engineers design statistical models, mimicking the required logic for prediction (e.g., logics used for recognising dialects). Using statistical models, engineers translate logics to computer understandable languages. These are sourced, from public repositories, or written from scratch (Vasilev et al., 2019). Using mostly open-source development suites (e.g., TensorFlow, PyTorch, and Keras) (Stancin & Jovic, 2019) statistical models are fed training data. Iterations of prediction models are generated, each randomly altering parameters to achieve better prediction accuracy. A different set of data entries are used to determine the accuracy of prediction. For a dialect recognition model testing data could consist additional audio fragments from the same people as in the training data. Alternatively, testing data might consist of audio fragments of people absent in training data.

Comparing this description with the definitions in Table 1, we find two main variances in IP between software and AI. First, software products were replaced with prediction models. These models are not simply a different from of software, but rather a novel type of digital product. Second, source code was exchanged for statistical models, as these models are considered the building blocks of AI products, akin to source code in software. Leading to the final list of AI-related IP written in Table 2.

Table 2

Comparison forms of software-related IP and AI-related IP

Definition in software	Definition in AI
Software products;	Prediction models;
Source code (full or partial);	Statistical models;
Databases;	Databases;
Distinct visually perceptible aesthetic features;	Distinct visually perceptible aesthetic features;
Distinct aesthetic features not contained in products (e.g., logo's);	Distinct aesthetic features not contained in products (e.g., logo's);
Undisclosed information receiving its commercial value by being secret.	Undisclosed information receiving its commercial value by being secret.

Note. From WTO. (1995). *Overview: the TRIPS Agreement*. https://www.wto.org/english/tratop_e/trips_e/intel2_e.htm#relatedright

2.3. Software and AI: differences in IP strategies

The comparison of IP provided two variances in IP assets. Blind et al. (2006) argued that such variances can be used investigate IP strategy by signalling relevant IP-assets and examining innovation-specific factors. Several relationships between innovation-specific factors and IP strategies are overviewed below and summarised in Table 3.

A case-study by Chatterjee (2017) case-study on AI SuSu's investments indicated that AI SuSu's spend most of their investments on R&D and data related assets. Also, Tantleff (2015) noted that the ability for AI SuSu's to generate additional revenues. Lastly, Wachowicz & Gonçalves (2019) indicated that AI SuSu's can exploit their databases for multiple clients and purposes. Overall, these studies suggest that, compared to software, data and related assets are increasingly seen as key resources in the field of AI. This shift has two implications. Firstly, copyrights become near-impossible to enforce as key assets are primarily written by non-human actors, who are ineligible for legal authorship (Kop, 2020a). Secondly, propensities for using database rights are expected to increase given its increase in value compared (Maurer et al., 2001).

Hall et al. (2014) reviewed economic literature dedicated to IP protection decision-making by software SuSu's and concluded IP strategies are affected by the degree of tacitness. Tacit knowledge relates to knowledge that is difficult to write down, based on experience of skills (Nonaka & von Krogh, 2009). Codified knowledge on the other hand related to knowledge that can easily be written down or visualised, based on facts and figures (Nonaka & von Krogh, 2009). Regarding AI, Wachowicz & Gonçalves (2019) argued that AI-related IP is comparatively more tacit by nature because data processing and statistical model design require substantially more personal judgement and expertise.

One resulting difference might be a negative impact on technological complexity. Samuelson & Scotchmer (2002) compared the usage of technological complexity between traditional manufacturing, semiconductor, and software. Software-related IP appeared more tacit compared to the former two industries, relying less on standardised processes and more on expertise. Making tacit assets technological complexity would imply intentionally complicating knowledge embedded in the minds of individuals, doing so was deemed near impossible as firms cannot manipulate the minds of employees (Samuelson & Scotchmer, 2002). This caused software SuSu's to become generally disinterested in technological complexity (Samuelson & Scotchmer, 2002). Based on this and considering AI-related IP increases in tacitness, we predict a greater disinterest in technological complexity.

Another difference was posited by James et al. (2013) who found software SuSu's were less inclined to utilise any legal instrument for IP protection because of the challenges involved with legally defining tacit knowledge. Based on the comparative increase in tacitness, AI SuSu's are expected less inclined to use formal mechanisms.

Kearns & Lederer (2003) studied the effects of knowledge sharing and competitive advantage in software SuSu's and found how the ambiguity of knowledge affected the design and success of IP strategies. Ambiguity raises difficulties in specifying knowledge and determining how knowledge led to certain outcomes (Law, 2014).

In the context of AI it is oftentimes not possible to see how models have reached their prediction (Fromer, 2019). Also, Fromer (2019) noted that it is usually not possible to determine how specific data entries or labels led to final predictions. Consequently, AI-related IP is considered more ambiguous compared to software-related IP.

Consequently, is greater reliance on secrecy. Software SuSu's have attributed their usage of secrecy due to greater degrees of ambiguity (Klein, 2020a). Ambiguity is regarded as a tool against infringement, trusted to sufficiently block imitation on its own (Klein, 2020a). Consequently firms disregard most appropriability mechanisms, depending solely on secrecy (Klein, 2020a).

Another consequence of this may be negative impact on technological complexity usage. This mechanism relies on the ability to control to the final design of end-products. AI SuSu's lack this ability since the disconnect between input and outputs components makes them incapable of predicting nor controlling end-product design. Technological complexity can thus not be applied for a majority of AI-related IP, suggesting a reduced usage of this mechanism.

On the other hand, increased ambiguity may cause AI SuSu's to develop ways of applying technological complexity, circumventing the limitations introduced by separated inputs and outputs. SuSu's in emerging fields have been shown to innovate around similar limitations, producing novel applications of technological complexity such as clone detection and AI-assisted obfuscation technologies (Canfora & Di Penta, 2007). Considering this, it is pertinent to remain sensitive to novel examples of technological complexity.

The development process overview revealed AI SuSu's to extensively use open-source assets, more so than software SuSu's. Mockus & Herbsleb (2002) demonstrated how pervasive usage of open-source assets caused Apache and Mozilla to adopt different IP strategies. These findings were supported by Samuelson (2010) who illustrated how the degree of open-source asset conditioned the way AI SuSu's protect and appropriate investments.

Increased usage of open-source tools has been shown to increase the propensity for software SuSu's to utilise complementary assets (Von Krogh et al., 2003). This noted by Von Krogh et al. (2003) primarily because open-source tools involve creative commons licenses, limiting the commercialisation of the product. Software SuSu's develop additional proprietary assets, not subject to creative commons licenses, generating returns via "freemium" or revenue sharing structures (Pujol, 2012). Considering open-source assets are used more extensively in AI, we predict this relationship to persist for AI SuSu's.

Kim (2007) investigated the effects of creative commons licenses on small software SuSu's, concluding that increases in open-source assets are usually associated with general disuse of formal mechanisms. This because legally protecting products, developed using open-source tools, is generally prohibited by these licenses (Kim, 2007). Accordingly, we predict AI SuSu's exhibit similar propensities for legal instrument usage.

A study on software SuSu's by Gyimóthy et al. (2005) found a negative relationship between open-source assets usage and the utilisation of technological complexity. Gyimóthy et al. (2005) concluded that open-source assets undermine the practicably of technological complexity, requiring massive changes before becoming unrecognisable by competitors, which is not cost-effective. Similar, if not stronger, tendencies are expected considering the even greater usage of open-source assets.

Von Hippel (2005) showed how software SuSu's, who used open-source tools, were likely to share open-source ideologies; free revealing of information and low-cost diffusion of ideas (Von Hippel, 2005). As a result, these software SuSu's made less use of secrecy as this would conflict with their open-source ideologies (Von Hippel, 2005). The expectation is that AI SuSu's hold these ideologies to an even higher degree, in turn causing the disuse of secrecy

Table 3

Innovation-specific factors and expected differences they can generate

Name of IP-related elements and sources	Expected difference(s) in IP strategy
Key resource shift (Blind et al., 2006; Chatterjee, 2017b; Tantleff, 2015a; M Wachowicz & Gonçalves, 2019)	More use of database rights, reduced copyright usage
Increased tacitness (Hall et al., 2014; Samuelson & Scotchmer, 2002; Marcos Wachowicz & Ruthes Gonçalves, 2019)	Less technological complexity, less legal instrument usage

Increased ambiguity (Canfora & Di Penta, 2007; Fromer, 2019; Kearns & Lederer, 2003)	More secrecy, novel obfuscation technologies
Use of open-source assets (Gyimóthy et al., 2005; Kim, 2007; Mockus & Herbsleb, 2000; von Hippel, 2005; Von Krogh et al., 2003)	More complementary assets, less legal instrument usage, less technological complexity, less secrecy

Looking beyond innovation-specific factors, two additional factors were found, written in Table 4. Other researchers looked towards market conditions IP strategy design. Jaworski & Kohli (1993) introduced a series of validated measures meant to assess markets titled "MARKOR". These measures pertained to market orientation, intelligence generation, intelligence dissemination, and the actions of marketing and nonmarketing informants. Doing so allowed the authors to explain why firms were more market-orientated than others and identify linkages between market conditions and business performance.

It has been shown that using only market orientation will suffice when studying innovation strategies in rapidly changing industries (Lynn & Akgün, 1998). The other measures are less telling, more concerned with marketing actions and knowledge development or diffusion (Lynn & Akgün, 1998). Market orientation relates to the alignment of firm strategy in accordance to market conditions (Lynn & Akgün, 1998).

Market orientation was extended during a study on the impact of lead-time usage by software SuSu's (Chen et al., 2005). Chen et al. (2005) dissected the concept into four market conditions: technological novelty, technological turbulence, market turbulence, and market newness. Our basic assumption was that differences in market conditions, between AI and software, could provide insight into areas in which IP strategies may differ. To this end, market conditions in each industry were compared to assess whether these could be used as factors to explain IP strategy discrepancies.

Technological novelty is determined by how new technologies, used to develop products, are considered by the firm (Lynn & Akgün, 1998). Technological turbulence relates to the rate at which those technologies change (Lynn & Akgün, 1998). Between AI and software technological advances are considered to follow identical trajectories, both reliant on the same advances in computing power and development software (Kulkarni & Padmanabham, 2017). To this end, is assumed that the novelty and rate of change in technologies will remain consistent.

Market turbulence is determined by the rate of change in consumer needs and the rate of change in distributions channels (Lynn & Akgün, 1998). Software consumers become increasingly interested in AI integration into existing products for Wi-Fi (Kerravala, 2019) and climate control (Bloomberg, 2020). Indicating consumer needs are overlapping between the two industries. Distribution channels related to ways products are delivered are similar in AI and software, dependent identical electronic technologies (e.g., computers, mobiles devices, etc.) (Lins et al., 2021). The above suggests consumer needs and distribution channels follow identical trajectories between the two industries, implying market turbulence is consistent. With respect to our basic assumption: only differing market conditions provide insights on IP strategies, market turbulence was omitted.

Market newness determined by the novelty of consumer needs and novelty of distribution channels (Lynn & Akgün, 1998). The market for AI SuSu's is inherently more novel than software SuSu, as only recent advances in computing power and software have lowered the barriers for entry, allowing SuSu's to enter the AI market (Delponte, 2018). Consequently, market newness is used to investigate market effects on IP strategies. Taipale (2010) used this factor to a case study on a Finnish software SuSu and found positive relationships between market newness and lead-time usage. Doing so ensured products corresponded with novel consumer needs and distribution channels as they arise (Taipale, 2010). The novelty of the market for AI SuSu's can thusly predict an increase in lead-time usage. Levine & Sichelman (2018) used, among others, market newness to investigate the reasons for or against the usage of trade secrets among software start-ups. Their conclusion was that software start-ups, embedding in more uncertain markets were more prone to use secrecy to protect IP. Consequently, the expectation is that, since the AI market is more novel, SuSu's will rely more on secrecy compared to software SuSu's.

Kochar (2020) noted that both software and AI development require focal knowledge, basic skill in writing code, designing interfaces, and testing products. However, AI development requires additional domain knowledge related to the use-case of products (Kochar, 2020). For instance, building a model meant to recognise cancerous cells requires additional research into the medical domain. Thus, according to Kochar (2020) R&D in AI development will typically take up larger shares of total investment, implying R&D intensities are higher in AI compared to software. R&D intensities have been shown to affect IP strategies in young highly innovative firms (Veugelers & Schneider, 2018). The authors found positive associations between high R&D intensities, using a threshold value of 73 per cent of total investment to distinguish between high and low R&D intensities, and the usage of secrecy. We therefore expect comparably more usage of secrecy by AI SuSu's.

Table 4

External factors and expected difference(s) in IP strategy

Name of external factor and source(s)	Expected difference(s) in IP strategy
Market newness (Chen et al., 2005; Levine & Sichelman, 2018; Taipale, 2010)	More lead-times
R&D intensity (Veugelers & Schneider, 2018)	More secrecy

3. Methods

The empirical approach is discussed and justified below, including data collection and analysis. Details on the sample and sampling process are provided and an overview is given on the process followed to translate preconceptions into measurable themes and sub-themes. Next, the way findings led to an answering the research question are discussed. Lastly, is section on research quality and initial reflection on the methodology.

3.1. Research design

The study was concerned with the influence of perceptions of various factors on IP strategies. It intended to explain these effects in a generalised manner, relevant across a large group of AI SuSu's. Consequently, the study followed a qualitative cross-sectional research design.

The similarities between AI and software suggested extent theory on software SuSu's could be used for the investigation of AI SuSu's. Consequently, an abductive approach posited by Timmermans & Tavory (2012) was used, to incorporate existing concepts while remaining sensitive to novel concepts.

Data were analysed thematically to determine the relationships between meanings (themes) and actions (Braun & Clarke, 2012). Initially introduced by Fereday & Muir-Cochraine (2006) and recently advanced by Braun & Clarke (2012), thematic analysis uses theory-driven themes, while also allowing for novel themes to emerge from data by combining deductive *a priori* code formulation by Crabtree & Miller (1999), with Boyatzis's (1998) process of inductively forming themes from raw data.

3.2. Data collection

Thematic analysis, following Fereday & Muir-Cochraine (2006), prefers several types of data collection including semi-structured interviews, observations, and content analysis. Two methods are typically combined, or the same method is used to collect data from different sources. For instance, an investigation on the opportunities of the digital age to early literacy relied on semi-structured interviews and observational data (Flewitt et al., 2015). Another study, on the manner in which consumers use health apps for self-care of chronic illnesses, relied wholly on semi-structured interviews from heterogenous stakeholders (Anderson et al., 2016).

The primarily source of data were semi-structured interviews centred around an interview guide, containing questions related to pre-existing theoretical concepts. Interviewees were however encouraged to provide their personal point of view, being free to bring up any unrelated concepts. Interviewers could explore these concepts further by asking follow-up question, departing the interview guide (Bryman, 2016). The sole requirement was that all questions in the guide were asked in each interview and that their phrasing remained consistent. Consistent phrasing between interviews safeguarded measurement reliability, ensuring participants interpreted questions in a similar fashion.

The added benefit of semi-structured interviews, compared to open interviews, is that it negates the risk of data overload (Chaiklin, 1991). Using the interview guide delineated the areas of investigation, ensuring data were collected on related factors (Timmermans & Tavory, 2012).

In accordance with Utrecht University policy, all participants were offered informed consent forms prior to interviewing. Participants were sampled based on specific criteria. One, being the CEO or manager in charge of IP strategy at an AI SuSu's. Two, being a member of the NL AI Coalition. This coalition, composed of various AI SuSu's, regional innovation hubs, specialised law firms, and other AI stakeholders, facilitates knowledge sharing and supports pilot projects and research on AI. Its members are of varying age and sub-fields and were more easily recruited due to the affiliation with the NL AI Coalition "SuSu working group". Additional data were gathered by attending network events and workshops, dealing with IP and IP-related topics, hosted by the coalition.

Initial recruitment commenced during tri-weekly networking events in which the study was introduced to AI SuSu's and stakeholders. Additionally, messages were sent in a community chat, to reach any absent AI SuSu's. Members were invited for interviews via email. Reminders were sent three days after initial invitation.

A total of nine respondents, spanning a broad range of market tenure, sub-fields, and disciplines, were recruited; interviews took between one and two hours. Seven AI respondents were SuSu CEOs (one with prior experience as an innovation hub manager). And two were digital IP advisors. AI incubator managers and IP advisors were recruited as these participants were able to verify whether AI SuSu's really acted in the manners indicated. Also, integrating these participants' perspectives facilitated the generalisation of findings as AI incubators possess experiences in managing multiple different AI SuSu's. See Appendix B for an overview of all invited AI SuSu's, including sub-field, and market tenure.

3.3. Operationalisation of preconceptions into themes

Before the interviews, preconceptions about the relationships between factors and IP strategies were translated into better manageable themes and sub-themes (Dubois & Gadde, 2002). Preconceptions were labelled, provided codes, defined, and descriptions were written for determining their presence. These were combined to form the code manual, given in Appendix A.

The four IP-related concepts were translated into the following key themes "Key resource shift" (Blind et al., 2006), "Increased tacitness" (Hall et al., 2014), "Use of open-source assets" (Samuelson, 2010), and "Increased ambiguity" (Kearns & Lederer, 2003).

Sub-themes for "Key resource shift" were based on underlying dimensions. "Copyright complexities" expected lesser interest in formal mechanisms as key resources shifted IP assets which are illegible for formal protection (Kop, 2020a). "Database right dominance" predicted greater utilisation of database rights due to the increased value of data (Maurer et al., 2001).

For "Increased tacitness" two sub-themes were added. "Legal definition difficulties" related to the negative impact on the usage of formal mechanisms (James et al., 2013b). And "Technologically complex tacitness" following Samuelson & Scotchmer (2002), who found tacitness to negatively impact technological complexity usage.

Four sub-themes were included for key-theme “Use of open-source assets”. “Creative commons prohibition” building on findings by Kim (2007) who found negative effects of creative commons licenses on the usage of all legal IP protection instruments. “Conflicting interests” suggested by Von Hippel (2005) who found the motivations associated with open-source asset usage to have a negative impact on the usage of formal mechanisms. Sub-theme “Open-source complementaries” based on Von Krogh et al., (2003) who found positive associations between increased open-source asset usage and complementary assets. Lastly, “Easy open-source reverse engineering” posited by Gyimóthy et al., (2005) who concluded software SuSu’s were less likely to depend on technological complexity when assets were mostly open source.

For key-theme “Increased ambiguity” two sub-themes were found. “Ambiguous secrecy” predicted the usage of secrecy to increase due to greater ambiguity of information (Klein, 2020b). “Novel obfuscation technologies” resulted from findings by Canfora & Di Penta (2007) who showed that novel obfuscation technologies emerged from software SuSu’s led to sub-theme “Novel obfuscation technologies”. This to capture novel application of obfuscation technologies.

Chen et al. (2005) found market newness useful for assessing IP strategies. To capture similar effects the key-theme “Market newness” was added. Sub-theme “Novel consumers” was added describing the relationship between novel consumers and lead-time usage based on findings by Taipale (2010). Another sub-theme titled “Novel distribution” was added following Levine & Sichelman (2018) who demonstrated a positive relationship between novel distribution channels and secrecy.

For “R&D intensity” sub-theme “Complex confidentiality” was added based on findings by Veugelers & Schneider (2018) who concluded high R&D-intense firms were more likely to depend on secrecy compared to firms with low-intensities. Veugelers & Schneider (2018) utilised a threshold value, which was carries over to this study, of 75 per cent to distinguish between the two types of firms.

The initial interview guide consisted of a general introduction, questions, and concluding formalities. Questions, written in Table 5 The interview guide was adapted continuously, adding questions to address emerging themes similar to Dubois & Gadde (2002).

Table 5

Interview guide consisting of main and elaboration questions, with respective codes and sources

, were divided into core and elaboration questions. Core questions were formulated as to not influence answer while directly addressing key-themes. Elaboration questions were formulated in a similar fashion but addressed sub-themes or meant to probe for novel themes. The interview guide was adapted continuously, adding questions to address emerging themes similar to Dubois & Gadde (2002).

Table 5

Interview guide consisting of main and elaboration questions, with respective codes and sources

Core question	Elaboration question	Theme and source(s)
Would you consider your market more dynamic than software due to the novelty of AI? *(Ex. Customer needs, competition, technologies, etc)*	How has this influenced your approach to protecting your IP? Would you say that you need to continually cater to novel customers? When doing so, are you also required to employ novel distribution channels? How has this played a role in your usage of IP protection?	Market newness (Chen et al., 2005; Levine & Sichelman, 2018; Taipale, 2010)
In what way do you stay ahead of competition?	Were there any methods that you find unsuitable because of this? (If not mentioned before) In what way has this affected your ability to keep IP secret from competitors?	Market newness (Chen et al., 2005; Levine & Sichelman, 2018; Taipale, 2010)
Which single resource or asset would you say is most important to the success of your product?	How did the strong reliance on data (or alternative answer to prior) impact IP protection? (If indicated data) Which legal protection instrument is most relevant to you? (If not mentioned) And copyrights and database rights?	Key resource shift (Blind et al., 2006; Chatterjee, 2017a; Maurer et al., 2001; Tantleff, 2015b; Marcos Wachowicz & Ruthes Gonçalves, 2019)
Would you say that R&D takes up more than 75 per cent of your total investments?	What implications did/does this emphasis on R&D (or alternative answer) have on the way you protect IP? (If indicated as high R&D intense) Is secrecy a viable option to protection the resulting IP? Why or why not?	R&D intensity (Veugelers & Schneider, 2018)
Is the knowledge needed for AI development easy or difficult to put into words? *(Ex: Are skills easily transferred among people, or does it take first-hand experience and "feeling"?)*	In which ways did this tacit (or alternative answer) nature of your IP influence your chosen IP strategy? If not mentioned) In what ways did/does this influence your usage of formal mechanisms? Was making your product more technologically complex an option to protect IP? Why or why not? (If not discussed) How did the nature of IP play a role in determining the viability of this option?	Increased ambiguity (Canfora & Di Penta, 2007; Fromer, 2019; Kearns & Lederer, 2003)

How big of a role do open-source assets play in the development of your product?	Which implications did this have on your method of protecting IP? (If not brought up) How did this affect your attitude towards protection IP via legal means? (If indicated big role) Do you adhere to open-source values because of this large role? And, has this deterred you from utilising forms of IP protection? (If unmentioned) Would keeping IP secret conflict with these motives? Did/do you consider exploiting the open-source community by letting third-party developers build additional features on top of your product? (If not mentioned before) Do/did you see this pervasive open-source usage as a threat to ease of reverse engineering your product?	Use of open-source assets *Comparison IP in software and AI,* (Gyimóthy et al., 2005; Kim, 2007; von Hippel, 2005; Von Hippel & Von Krogh, 2003)
Do you feel that competitors will have difficulties in determining your data processing and statistical model design?	Why or why? How did knowing this (or alternative answer to prior question) affect decisions related to IP protection? In which ways did you design your product to complicate reverse engineering? (If not brought up) How did the separation of input and output influence this process?	Increased ambiguity *Comparison IP in software and AI,* (Canfora & Di Penta, 2007)

3.4. Data analysis

First audio recordings of interviews were transcribed using Vosk (2021), an AI speech recognition toolkit developed by Alphacephei. The pre-trained models Kaldi_NL (2021) and Aspire 2.0 (2021) were used for Dutch and English interviews respectively. The raw output was revised by hand, altering the lay-out, and redacting sensitive pieces of information.

Finalised transcripts were imported into NVivo, release 1.5 (for MacOS) (QSR International Pty Ltd., 2020). Next the code manual was imported into NVivo (QSR International Pty Ltd., 2020), adding key themes and sub-themes as nodes and sub-nodes respectively.

Each transcript was open coded individually to generate a broad selection of codes and get better acquainted with the data. Pieces of data were analysed line-for-line. Pieces of data, relating to existing codes, were categorised accordingly. Nonmatching pieces of data were given novel codes and grouped under the placeholder key-theme “emergent codes” (Boyatzis, 1998).

The large number of codes described small aspects of IP strategies, lacking explanatory power. This was addressed by merging similar codes, forming codes which were applicable across multiple cases. Codes appearing to be a component of another code were inserted underneath the primary code as sub-codes, forming child- and parent-code structures. Remaining codes only appearing in single cases were considered irrelevant, lacking explanatory power, and subsequently removed.

Findings needed to reveal the relationships between themes, as the research question called for an understanding of the influence of innovation market concepts on IP strategy design. To this end, it was necessary to also identify the interactions between themes. Relationships were drawn between codes, recognising synergies or inconsistencies.

The research question was answered by combining three types of findings. First, findings confirming the presence of theory-driven themes exemplify similarities between IP strategies. Second, results suggesting theory-driven themes must be adapted indicate areas in which IP strategies differ between AI and software. And third, emergent data-driven themes illustrate previously unobserved concepts leading to IP strategy differences.

3.5. Research quality and reflection

Contributing to the quality of research were several measures. Collating transcripts enables other research to replicate the approach and derive similar results, supporting replicability (Chaiklin, 1991). Predetermining definitions and descriptions of themes ensured measurement remained consistent between cases, buttressing internal reliability. Coding followed several iterations, shifting between code manual and (older) data. Doing so, findings additional evidence themes which were found in later interviews. This ensured all causal relationships were identified and further validated findings based on additional evidence, found in previous interviews. Codes were consolidated to form overarching codes applicable across multiple cases. Doing so ensured findings remained applicable to the greater population of AI SuSu's, buttressing external validity. Pigeonholing can occur when existing theories are incorporated too prematurely, causing data to be incorrectly forced into existing codes. The code manual was disregarded during initial rounds of coding, to ensure initial codes remained unaffected. Data saturation is required to ensure no findings are overlooked. This was witnessed once analysis of one additional interview did not lead to an extension or creation of (novel) codes (Chaiklin, 1991). Initial findings were discussed with AI stakeholders throughout to ensure data were correctly interpretated, supporting internal reliability. Finally, direct quotations were used to present results, demonstrating the robustness of findings.

However, despite these measures some methodological shortcomings were acknowledged. The lack of longitudinal data limits the ability to truly measure cause and effect relationships. In other words, IP strategies might have caused influenced innovation-specific or external factors. Given the novelty of AI, open interviews could be more valuable, being more accommodating for entirely new factors to emerge. Data were collected solely through semi-structured interviews, threatening internal reliability as findings could not be cross-referenced. For example, observation data, gathered by being present at meetings at AI SuSu's could have triangulated interview data to verify whether concerns voiced in interviews emerged in real-life. Or, textual analysis, of AI SuSu documents on IP strategies, would have facilitated data triangulation by providing a different source of "black and white" data. However, COVID-19 restrictions ruled out the ability for safe company visits. Participants were hesitant to provide access to business documents on their IP strategy, negating textual analysis. This threat to internal reliability was however reduced by combining interview data from various sources to verify whether strategies discussed in interviews were truly adopted in practice. Thematic analysis has been shown to produce inconsistent findings, due to its flexible nature (Holloway & Todres, 2003), this was mitigated by continuously discussing findings and interpretations with respondent and other stakeholders.

4. Results

Results are presented per theme, each section closing with a series of key findings (See tables 6 through 11).

First and foremost, based on a comment by respondent 3, AI SuSu's "do not distinguish between IT and AI [...] because you are always looking at both." Respondent 6's comment reinforced this: "Firstly, taking a step back, we are a software company. The software we develop has two facets. One is the software […] enabling us to build the models and AI. […] And the other is data security, because showing that you are in control [of your data] is extremely important." Respondent 2 clarified as "[f]or every little part of the AI, there is a huge part of the software that you have to make." These findings imply AI and software overlap, corroborating conclusions derived from the comparison IP between AI and software and validating the abductive approach.

4.1. Key resource shift

Theory-driven theme "Key resource shift" received support from all studied SuSu's. Respondent 3 pointed out that: "about […] anything to 60 to 80 per cent of your AI project is on data. It's just getting the data correct. Planning it, labelling it, formatting it, putting it in the right [category]." Respondent 4 elaborated: "our IP is in our data and the way our 'pipelines' work […] The 'magic sauce' is the way we bring everything together […] It's how we make data, label data, and train models".

Indeed, confirming the theme by illustrating the importance of data and related processes. However, also indicating that key resources include data collection and the ways in which AI products are distributed to consumers.

Sub-theme "Copyright complexities" is exemplified by respondent 8, who "would not have a clue how to get copyrights on our model, except maybe out back-end." Respondent 5 was equally unsure about copyrights, going so far as to commission "[a] study [...] (on copyrights in AI) […] that concluded that AI-related IP cannot be protected (using copyrights), because it's mostly […] software."

Instead, sub-theme "Database right dominance" expected AI SuSu's to increasingly depend on database rights. However, findings suggested otherwise, for a variety of reasons.

Some were simply "not aware of database rights", such as respondent 8. This was explained as they "made a model based on a 'convolutional neural network' […] Large databases are less important for these types of models […] [I]t's a different way of approaching AI" Thus suggesting that use-case and characteristics of statistical models may reduce interest in database rights.

While others, like respondent 5 stated: "We are provi[ded] […] with the data […] need[ed] to create the model […] and that data is always owned by the client." Respondent 8 concurred: "The end-user is the sole owner of the images. We use those images to train our model. We receive that component, having our model interpret the images, free of charge. Actually, clients pay us to interpret their images. Those images are on our servers but will never be our property." Implying that clients, insisting against the transfer of data ownership, prohibit the usage of database rights.

Respondent 1 mentioned to "not fully understand (database rights) […] In one law case where a database was comprised of satellite data […] effort was insufficient to receive database rights. While in another law case, […] involving physical transportation of data […] effort was sufficient." Additionally, respondent 9 explained how the importance of data, led them to "includ[e] them (database rights) in our terms and conditions [...]." However, that these had to be "written by a specialised [data base rights] attorney [before] […] things became clearer." This suggests that unclear descriptions of the requirements, needed to receive database rights, dissuade AI SuSu's from utilising database rights. AI SuSu's abstain from its usage since they are unsure whether effort is deemed sufficient. Consequently, database rights are used only after AI SuSu's have consulted with costly specialised law consultants.

Another consequence of shifting key resources, was introduced by respondent 6 who was "currently busy making deals with clients to build a 'mega database'." Respondent 2 held similar motives "retrieve[ing] large amount of data [...] [by] scraping (and storing) a lot of data from the internet." For the "main reason […] that […] when you want to re-train. […], [y]ou always want to test and re-test […] with the same documents." Respondent 4 would "prefer to store all data indefinitely. However, that is just not possible […] [because] it would be financially and legally unfeasible."

The above led to the data-driven theme "Mega databases", describing the continuous expansion of databases. By doing so AI SuSu's generate a valuable piece of IP useful in several ways. One, large amounts of archived data are used for developing other AI products, for related use-cases. Two, archived data is used to test old models against new models, comparing the accuracy of prediction. Three, new additions of the database are used to re-train existing models to improve performance. However, this strategy is not practical in all scenarios. For certain clients who prohibit the storage of data, requiring all data to be removed once products have been delivered. Also, for lesser-resourced firms because of the high costs involved with data collection and storage.

Table 6
Key findings for theory and data-driven themes associated with "Key resource shift"

Theme or sub-themes (in italics)	Key findings
Key resource shift	Key resources shift towards data, data processing. Also, data collection and manner which AI products are delivered to end-users becomes more relevant. Respondents call these resources the "data pipeline."
Copyright complexities	AI SuSu's consider AI-related IP to be illegible for legal authorship, foregoing the usage of copyrights.

Database right dominance	Database rights are used sparsely, due to unclear descriptions of conditions to receive database rights, client requirements, and because of unfamiliarity with this instrument.
*Mega databases**	Shifts towards the “data pipeline” ensued a strategy to build IP by continuously integrating novel data to build a “mega database” used for re-training or testing different models.

*Note. *Data-driven theme*

4.2. Use of open-source assets

The second key theme on open-source asset usage is exemplified by respondent 8: “Yes, open-source, we wouldn’t be here without it” and respondent 2. “Yeah, thank Google for open sourcing a lot (of assets).” Indeed, confirming AI SuSu’s make pervasive usage of open-source assets.

Results in relation to sub-theme “Conflicting mentalities” include respondent 9 who “contribute[s] to the (open-source) libraries [and] make it a priority [to] only use open-source assets but remain active in the (open-source) community.” Respondent 2 concurred, being “very open for sharing knowledge and everything (related to their product).” The same respondent explained that because of this they “personally do not like [...] secrecy [...] [We would tell] how it works in general concepts of the AI […] from which someone else can take it and implement it themselves.” Findings thus confirm AI SuSu’s hold open-source mentalities, incentivised to not only make use of open-source communities but also make contributions to those same communities. As a result, AI SuSu’s are more interested in share knowledge about products with others rather than keeping it secret. Moreover, the willingness to do so increases when the other party has displayed similar open-source beliefs, offering information of its own.

Respondent 8’s comment summarises evidence for sub-theme “Creative commons prohibition” remarking: “If you look at our front-end, the viewer in which customers receive the data, that’s all pretty standard (open-source assets) [...] Except our back-end. But our back-end uses a lot of open-source assets as well (inhibiting us from using formal mechanisms).”

Most assets used for AI development do indeed involve creative commons licenses, prohibiting copyrighting or otherwise formally protecting resulting products. Consequently, these limitations are attributed to disuse of form mechanisms, validating sub-theme “Creative commons prohibition”.

In terms of sub-theme “Open-source complementaries”, respondents did not mention whether open-source assets affected their propensity to utilise complementary assets. Instead, respondent 5 put forth two alternate factors negatively affecting complementary asset usage: “simply we lack the […] momentum […] both in terms of product and market share.”, suggesting that complementary asset usage is negatively influenced as AI SuSu’s believe their products will not attract a large community of third-party developers because of lacking performance and/or consumer-base.

Next, findings on “Easy open-source reverse engineering” are best captured by respondent 2’s comment: “The paper […] has kind of the details on the experiments you made [and] all the details on how the approach works. […] [G]iven this, implementing something like this in AI is very easy once you understand it and read the paper.” Proposing that in general AI products, making pervasive use of open-source, are easily copied. However, we are unable to confirm or deny this expectation because no respondents indicated how this influenced their usage of technological complexity, we are unable to confirm or deny this expectation.

From the data emerged sub-theme “Accidental infringement” as several respondents indicated how extensive usage of open-source assets could lead to accidental IP infringement. Accidental IP infringement is not directly related to IP strategies. Despite this we still consider it noteworthy to discuss as the consequences of doing seriously threaten firms.

Respondent 1 provided two examples of accidental infringement: “If you look at open-source, a lot of people think ‘I am working open-source, so I will never have to deal with IP’, that is simply not true […] And another thing is, products are usually open-source, because they are built on an (open-source) platform or application […] just because you make your product open-source, does not get you off scot-free.”

When asked whether creative commons licenses were checked respondent 3 replied “If I am honest. I would say no.” Additionally, the respondent provided two reasons why creative commons licenses are overlooked as they “do not have the resources” and “because our liability is so small […] they (the other firm) have no clue.”

Respondent 7 emphasised the risks of overlooking creative commons licenses: “It (issues with creative commons licenses) might destroy your business model. [I]f you think something is open-source, investigate what the underlying agreements mean in relation to your business model” and offered advice: “First, make sure your contracts are in order. Also, make sure that the agreements you made with the provider of open-source assets are clear.”

In sum, AI SuSu’s assume infringement is possible when freely publishing their product. Additionally, the implications of using creative commons are not carefully considered for lack of resources or because AI SuSu’s believe they will not be caught. Consequently, AI SuSu’s place greater importance on contracts and agreements, stipulating who owns resulting IP, to which degree open-source assets can be modified, and how resulting IP can be commercialised.

Data-driven Sub-theme “Third-party service usage” emerged as all cases mentioned that AI development depended heavily on third-party services, provided by large corporations. Chiefly are “… databases, cloud storage, and virtual machine. And […] Azure machine learning […], to train our models”, according to respondent 6. Respondent 2 stated “buy[ing] your own machine […] is very expensive. [B]asically if you are a start-up. In AI, it’s kind of infeasible (without using third-party services).”

Respondent 6 identified this as “a big risk in this. If, for example, you have based your entire product on their service. Because when Amazon raises their prices, your margins start to shrink.” They elaborated, explaining its impact on general strategy because “[i]f Amazon decides to raise their prices by 30 euros, your entire business plan may collapse.”

Overall, AI SuSu’s cannot develop products without depending on third-party services. Because the alternative, purchasing storage and machine learning hardware, is too costly. Heavy reliance on these services threatens profitability because providers can simply raise usage fees. To this note, it is expected that third-party service dependencies have implications on IP strategies in the context of AI.

Table 7
Key findings for theory- and data-driven themes associated with “Use of open-source assets”

Theme or sub-theme (in italics)	Key findings
Use of open-source assets	AI SuSu’s utilise open-source assets throughout product development.
Conflicting mentalities	As consequence of extensive usage open-source assets AI SuSu’s hold open-source attitudes, deterring the usage of secrecy. Information is freely shared, with like-minded individuals, and emphasis is placed on contributing to open-source communities.
Creative commons prohibition	Indeed, barriers erected by CC licenses cause AI SuSu’s to disregard formal mechanisms.
Open-source complementaries	AI SuSu’s do not depend on complementary assets because of quantity of material; data and model performance and market share were considered insufficient for attracting large third-party developer communities.
Easy open-source reverse engineering	Findings were inconclusive regarding this theme.
*Accidental infringement**	Some AI SuSu’s, utilising open-source assets accidentally infringe on IP, for lack of resources or by disregard for getting caught infringing. Because of this AI SuSu’s are more likely to utilise semi-formal mechanisms, utilising contracts with open-source providers, clearly delineating what can and cannot be done with resulting products.
*Third-party services**	AI development relies heavily on third-party services to host models, store data, and train models. This reliance threatens profitability and might also have implications for IP strategies.

*Note. *Data-driven theme.*

4.3. Increased tacitness

The key theme “Increased tacitness”, expected AI-related IP to be increasingly tacit compared to software-related IP and is verified in data by respondent 8 who illustrated the difficulties they experienced when codifying and diffusing knowledge between data scientists: “Yeah all credits go to <redacted> (respondent’s partner), because he gathers all the tacit knowledge. […] Because that’s incredibly difficult […], especially dispersing that tacit knowledge amongst all your people. […] I still have not found a proper way to do that.”

Findings on sub-theme “Legal definition difficulties” were consistent across all cases, exemplified by respondent 5: “I understand that it is almost not possible to protect anything without it being coupled to a physical product. Neither in Europe of America. So, we gave up on that.”

In sum, AI SuSu’s unable to codify why and for which reasons certain decisions were made. For instance, clarifying how data were interpreted or describing the data labelling process. This makes it difficult for AI SuSu’s to capture their IP in legal terms, leading to disinterest in formal mechanisms.

Respondent 3 offered another reason why tacit knowledge affected the usage of formal mechanisms: “Because […] IP, once you have it […] ‘How do you know other people are using it?’ and “How do defend it?’ […] As a start-up, you cannot, […] that is the reality of playing in this field”, implying AI SuSu’s go without formal mechanisms based on the believe that detecting AI-related IP infringement is unfeasible.

Table 8
Key findings for theory- and data-driven themes associated with “Increased tacitness”

Theme or sub-theme (in italics)	Key findings
Increased tacitness	AI-related knowledge is increasingly tacit by nature, described as the “art” of data science, complicating the diffusion of knowledge among employees.
Legal definition difficulties	AI SuSu’s experience difficulties in capturing AI-related IP in legal terms. Additionally, as knowledge becomes more tacit, AI SuSu’s find it challenging to catch competitors infringing IP. The outcome is a reduced utilisation of all legal instrument.

4.4. Increased ambiguity

Respondent 5 addressed key-theme “Increased ambiguity” commenting that “[d]etecting whether competitors have copied your IP is incredibly difficult because AI is so opaque compared to software.”

Respondent 6’s description of their sales tactics provided another insight: “Often we will say ‘give us a piece of your data, so we can throw it into the model’. We are then able to show whether it (the prediction) is correct with the client’s expectation. This was quite frightening in the beginning. You will always think will it truly be correct?”

AI SuSu’s find it challenging to detect IP infringement as input components are not easily observed. What data were used, how they were processed, and specific statistical models remain indefinable from an outside perspective. This they feel, differs compared to software products since these contain all pieces of IP required for its operation. Additionally, AI SuSu’s expressed difficulties in pinpointing which parameters or data entries led to an incorrect prediction. This exemplifying ambiguity from an internal perspective, where AI developers cannot explain how input components led to certain outcomes. This AI SuSu’s agreed differs for faulty software products because incorrect lines of source code are typically automatically flagged, allowing software developers to quickly spot what went wrong. Derived from these findings key-theme “Increased ambiguity” is indeed confirmed.

Sub-theme “Ambiguous secrecy” was supported in data. For instance, respondent 8 who described their strategy as “not tell[ing] anything and making sure people cannot access it.” This respondent attributed this to fact that other companies “can simply not access any of [the] back-end.” Additionally, when asked what the implications of the opaqueness of AI were respondent 5 mentioned that it allowed them to “just keep it (IP) secret.”

AI SuSu’s perceive ambiguity in terms of how easily end-users may access back-end assets. To this note, hosting models on in-house servers, offering products as a service, and edge-computing were seen as measures determining ambiguity. AI SuSu’s combine these measures with secrecy, trusting that this will provide sufficient protection against infringement.

Like findings for sub-theme “Easy open-source reverse engineering”, data on sub-theme “Ambiguous inability” proved insufficient because no respondent could specify whether increased ambiguity affected the propensity of utilising technological complexity.

Initial findings for sub-theme “Easy open-source reverse engineering” and “Ambiguous inability” remained inconclusive because respondent could not specify why technological complexity saw little usage.

However, insights related to “Novel obfuscation technologies” revealed a different picture, in which technological complexity was used except for other IP assets.

First a comment by respondent 2 illustrated how their focus went towards “[…] whole […] structure around it. The whole part of deployment, building a server with back-end.”

Signifying AI SuSu's utilise technological complexity but for assets other than finalised AI models. The interview guide was adapted accordingly letting us explore obfuscation technologies for a wider selection of assets.

Respondent 6 protected their datasets against unauthorised usage by "includ[ing] all kinds of restrictions in our data when working with partners […] written in contracts, in which we also mention samples. [F]or instance, micro-pixels that we include in photos, objects, or maps. […] We can trace those back when doing audits."

The same respondent provided another technique where "incorrect data […] are hidden within the dataset [unique] per partner. […] Datasets sold (without our permission) can then be easily traced back to individual partners."

Respondent 8 protected their web-portal, used by clients to interact with the AI model, against reverse engineering because "[y]ou can easily see which Django and Python […] packages are used." They would "scramble those, […] add[ing] fake references to packages which are not used for the product. […] This may impact your product's performance. A solution to this is to only do it for certain sections."

Further investigation uncovered several novel obfuscation technologies applied in AI. AI SuSu's employ measures to threaten litigation and simply auditing by including micro-pixels and/or fake data entries. The details and consequences of infringement are written contracts signed with partners, ensuring IP is sufficiently protected. Also, irrelevant references to statistical model libraries are added to products, which in fact are not required for the model to operate. Doing so, makes it more difficult for competitors to understand the innerworkings of products thereby protecting against reverse engineering. Each irrelevant package negatively impacts product performance, increasing load-times. This is mitigated by selectively adding packages to the most crucial assets.

Data-driven sub-theme "Partial disclosure" came forth from respondent 6 who would often have to prove "they are truly doing something with AI and are not like the others (companies developing non-AI products, using the term for marketing purposes only)." Because some clients or potential partner simply "do not understand it". Respondent 6 went on: "That (process leading up to prediction) the secret sauce. […] And you want to keep that close to the chest." However, "if I do that, not explaining how my model came to a prediction, people will ask me 'how do I know if it's correct?'."
Respondent 6 mentioned to "not tell everything. We do not explain how certain pieces of data lead to certain predictions [...] We do provide an overview of the entire dataset that was used to train the model. That's kind of an 'information overload' [...] you are kind of telling them too much." Respondent 3 backed this: "You'll find out […] very quickly. As much as you want to share what your product and your project or your IP is. […] [Y]ou need to just share a little bit."

The above led to the data-driven theme "Partial disclosure", describing the positive effects of ambiguity on partial disclosure. AI stakeholders lack sufficient understanding due to the ambiguity of AI-related knowledge. These stakeholders question whether AI was truly used. Or want to understand the process leading up to the prediction. This in turn, necessitates AI

SuSu's to disclosure information, which they would rather keep secret. IP is protected by partially disclosing IP, being mindful to not discuss any propriety assets. AI SuSu's will comprehensively explain products from a top-level perspective. Details on the type of data and how they were labelled are freely discussed. Also, AI SuSu's will freely discuss data processing, including data sources, what data were omitted, etc. However, AI SuSu's remain purposely vague on essential pieces of information such the individual relationships between data entries and final predictions. This comprehensive explanation provide stakeholders will too much information, convincing them that AI was truly used while not giving away essential pieces of information.

Table 9

*Key findings for theory- and data-driven themes associated with "Increased ambi*guity"

Theme or sub-theme (in italics)	Key findings
Increased ambiguity	AI SuSu's themselves often find it difficult to explain how data led to predictions.
Ambiguous secrecy	Higher degrees of ambiguity make detecting infringement difficult. Because of this AI SuSu's place greater importance on keeping IP secret.
Novel obfuscation technologies	Technological complexity is not applied on AI models, but instead on the data pipeline. AI SuSu's hide false data entries and pixels to simplify infringement audits. Also, false references to packages are added, defending against reverse engineering.
*Partial disclosure**	AI SuSu's are required to disclose information during negotiations to address the ambiguity of their IP. Simultaneously, AI SuSu's take advantage of increased ambiguity, overwhelming clients of partners with information.

*Note. *Data-driven theme*

4.5. Market newness

"Market newness" predicted that the market for AI SuSu's is comparably newer than software and that impacted IP strategies.

Respondent 4's comment illustrated this as "AI is still in its early days phase. A kind of cowboy western world, like we have seen for the internet in the 90s." Indeed, confirming the comparative market novelty of AI.

Respondent 2 was in support of the above stating: "While if you ask for AI in general how it's changing. It changes dramatically year by year. For example, the way I was doing the same task before my masters, like three years ago, and now after my master's it's completely different. Now it's much more automated, it's much easier. Before it could take me a week, now it takes me one day." They explained how they "had to just completely drop something […] To use something new, because in six months everything changed” and that because of this their current IP strategy was “almost zero. Which means we do not […] [W]e currently just do not have an explicit will to protect it.”

These findings relate less to market newness and more to market turbulence. Rapid advancements quickly made products obsolete. As a result, AI SuSu’s see no reason for protecting these assets.

“Novel consumers” related to an increased occurrence of lead-times due to the greater novelty of consumers. Respondent 6 explained how AI models are “quite dynamic and must be really flexible. Models need to be adapted after every small change in the (AI) landscape. Software products are usually one-size-fits-all. […] That is the major difference between (AI) models and software, you have to remain flexible and continuously tweak and adjust your product.”

The novelty of consumer needs requires AI SuSu’s to continuously adapt innovations. This differs in comparison to software, where products are put to market, typically not receiving major updates once launched. From this it can determined that indeed the novelty of AI stimulates the usage of lead-time.

Similarly, to the prior sub-theme, “Novel distribution” expected an increased usage of lead-times due to novel distribution channels. Most cases deliver products via cloud services and generated revenue via software-as-service approaches such as respondent 2 who “[is] [...] distributing our software as a service. […]. Once you have implemented [it], you can sell it any number of times […] and what we sell is access to the platform.” Respondent 9 utilises proprietary devices upon which the AI model “is programmed […] They call this “edge-computing.” Both examples are methods previously used in software, negating the expected novelty distribution channels. Additionally, no implications of this were revealed in the data.

Driven by data, was sub-theme “Developing legislation”, best illustrated by respondent 1 who illustrated how developing legislations affects the propensity for using formal approachability mechanisms: “There is a new category; artificial intelligence, which has emerged in the past couple of years within the field of software (IP). And it appears that this emerging area is leading to additional requirements (to utilise formal appropriability mechanisms). […] For instance, additional content-specific requirements emerge [...] One of which is that inventions must be described, so they are workable by others […] This was almost impossible to do[.]”

Related to findings on sub-theme “Database right dominance” which suggested database rights saw limited usage due to unclear definitions of ‘effort’. AI SuSu’s attribute this to the notion that these definitions are simply out-dated and have yet reached consensus on what can be considered IP or sufficient ‘effort’. Additionally, crucial topics related to AI such as the eligibility of synthetic or derivate databases have yet been included in legislation. AI SuSu’s therefore distrust formal mechanisms, negatively impacting the propensity to utilise formal mechanisms.

Table 10

Key findings for theory and data-driven themes associated with “Market newness”

Theme and sub-theme (in italics)	Key findings
Market newness	The AI market deemed newer and more dynamic compared to the software markets. Accounting for market dynamics, AI SuSu’s keep IP strategies to a minimum to not get locked into activities that could quickly become out-dated.
Novel consumers	Currently each AI use-case; client/consumer group, requires slightly different requirements, necessitating continuously redeveloped of AI models. Lead-times are therefore much more prevalent in the context of AI SuSu’s compared their software counterparts.
Novel distribution	Distributions channels, in the context of AI SuSu’s, are not more novel that those used by software SuSu’s. Thus, the novelty of distribution channels can be disregarded as a factor incapable of explaining IP strategies in this context.
*Developing legislation**	AI SuSu’s are hesitant to depend on formal mechanisms, such as database rights, because legislation has yet to reach consensus on what is considered IP or sufficient ‘effort’.

*Note. *Data-driven theme.*

4.6. R&D intensity

Findings related to "R&D intensity" which expected higher R&D intensities to raise the usage of secrecy, are presented below.

In general, most cases provided supportive evidence for key-theme. Respondent 4 for example noted "that 75 per cent of investments towards R&D [...] This may change in the future, but this is the case currently […] Yes indeed, you need domain knowledge." Respondent 6 confirmed to invest more than 75 per cent in R&D, wishing not to disclosure absolute figure. Lastly, respondent 8 stated "If you look at last year, then we generated about <redacted> euros in revenue. Of that revenue about <redacted> euros went towards R&D. And those are the salaries we pay our guys […] It's mostly the hours of people we hire, sometimes for a week or two."

AI SuSu's invest heavily in the development of domain knowledge by hiring experts or consultants. This to get a better grasp the specific use-case but also to determine their logic, understanding how these individuals came to certain conclusion or decisions. This knowledge is then used to build statistical models, mimicking this logic. Additionally, AI SuSu's dedicate substantial resources towards scanning for new models released by other firms. Staying up to date with the most cutting-edge models is a time-consuming process, primarily due to sheer quantities newly available models. Related and similarity resource-intensive is the testing of these new models. AI SuSu's run the existing and model in parallel, using the same data. The accuracies of prediction are compared to one another. New models exceeding current performance exchanged for the existing model.

However, several respondents indicated R&D intensities below 75 per cent. For example, respondent 2 who stated: "So, R&D (investments) are probably 10 to 20 per cent [...] It would be less, but we have two interns doing R&D." The respondent explained stating: "We have plans [...] to introduce new parts of the AI that are really [...] new. But that is more of a plan for the future, we are very early stage."

Or respondent 5 who stated: "I think, for us, investments are split 30 per cent towards R&D and 70 per cent towards market validation. This is done, in part, on purpose as we want to see whether people are interested and willing to pay money for our product before we make any substantial investments."

A final example came from respondent 9 who had downscaled R&D investments, below 75 per cent, to focus on "gathering as much material as possible […] and by doing so maintain your advantage over others."

The above suggest contradictory findings, negating us from confirming nor denying this sub-theme. However, data did propose initial explanations for these contradictories. Early-stage respondents, still in the process of finding fit-to-market were less inclined to invest in R&D for fear of developing products that do not optimally cater to user needs. Whereas more mature SuSu's were able to reduce R&D investments, focussing on support, sales, and marketing. Thus, partially explaining why R&D intensities differed among AI SuSu's.

In terms of IP strategy sub-theme "Complex confidentiality" predicting an increase in secrecy because of greater R&D intensities, was invalidated by data exemplified by low R&D intense respondent 5 who depended primary on secrecy as they "… recently got advised to keep everything that is invisible (to competitors) secret and to protect everything that is visible by other means [...] So, our strategy is to just keep it a secret."

Additional counter evidence came from low R&D intense respondent 9: "[O]ur primary objective is to remain legal owner of all those components (used to develop the product). Despite outsourcing our hardware development and assembly […] We do this with contracts. That's our primary focus, so there's never a discussion about ownership."

A response by respondent 4 to this sub-theme provided two other factors the usage of secrecy. One, because they intended to "use models for other purposes, as that is our IP and has to remain secret." And two, the notion that "[s]ome customers do not want the models to run on the internet […] for privacy reasons"

Freely sharing assets limits the ability to use those assets for other clients because past clients might litigate claiming ownership of the model or data.

Also, AI SuSu's fear missing out on potential revenue when revealing models to clients since these freely shared models are easily reverse engineered. Clients can then build additional products in-house, which would otherwise have been built by the AI SuSu's. The AI SuSu's thereby miss out on revenues that would have been generated from selling these additional models to the client. Also, AI SuSu's utilise secrecy because clients require data and models to be processed locally, usually remaining within European geographical boundaries. To this end, assets must be located on servers located at AI SuSu's.

Table 11

Key findings for theory and data-driven themes associated with "R&D intensity"

Theme or sub-theme (in italics)	Key findings
R&D intensity	AI SuSu's invested displayed varying degrees of R&D intensities, prohibiting confirmation or invalidation.
Complex confidentiality	Contradictory findings inhibit the identification of key findings regarding this sub-theme.

5. Conclusion

Designing and implementing an IP strategy has proven to be a primary tool in the arsenal of SuSu's. A well-designed IP strategy is relevant to SuSu's for two primary reasons. One, it ensures the limited resources held by SuSu's are optimally used. Two, designing a suitable IP strategy offers SuSu's more control over their situation, tackling any inherent uncertainties. Uncertainties have been shown to affect technological innovation and typically arise in emerging markets in which technologies, business models, and legislation are still developing.

One such uncertain emerging market is AI, evident from a recent survey that showed, despite recent technological advances, AI SuSu development has yet to accelerate due to concerns about AI-related IP issues such as ownership and legality (Delponte, 2018).

SuSu's are often associated with tackling societal challenges by introducing disruptive solutions and business models (Bradley et al., 2021). To this end, stimulating AI development by SuSu's becomes relevant as AI has been linked to tackling climate change and reducing poverty (Bradley et al., 2021).

Despite this potential and the desire for greater certainty, there appeared to be a dearth of research in the IP strategies utilised by AI SuSu's. Filling this gap provides a foundation from which IP protection can be studied in the context of AI. Also, existing IP strategy knowledge is extended by applying existing theories to novel empirical data. These initial findings led to the following research question:

"Which intellectual property strategies are used by start- and scale-ups in the field of artificial intelligence products?"

Wherein IP strategies were determined by the configuration of various appropriability mechanisms.

Studies by Wan et al. (2020) and Kulkarni & Padmanabham (2017) found software SuSu's to share the most similarities with AI SuSu's. However, an overview of software-related IP and AI-related IP put forth a series of factors, previously found to affect IP strategies, that differ between AI and software. An abductive approach was therefore employed, exploiting extent literature while remaining sensitive to novel insights.

First, literature on software SuSu IP strategies were reviewed to identify relevant appropriability mechanisms used by software SuSu's. Followed was literature review, guided by Teece (1984) and Hemphill (2004) who illustrated the effects of innovation and market factors on IP strategies. Six theory-driven themes emerged predicting IP strategy differences. Next, a variety of AI SuSu stakeholders were approached for semi-structure interviews in which factors identified in past literature and (novel) emergent factors could be addressed. Lastly, thematic analysis followed to find evidence for theory-driven themes and generate (novel) data-driven themes.

Results suggest a general increase in propensities for using informal appropriability mechanisms compared to their software SuSu's. Secrecy is utilised most as AI SuSu's trust in the protection provided from the ambiguity of AI-related IP. Also, models are kept confidential because free disclosure with clients may possibly restrict additional revenue. Additionally, the usage of secrecy is attributed to client requirements, demanding data and models are stored locally only. However, open-source motivations reduce propensities for secrecy. Rapidly changing consumer needs make lead-time more conducive, involving the continuous development of AI products.

AI SuSu's make limited use of complementary assets, not expecting to attract large open-source development communities due to lacking material and market share. Products developed by SuSu's as typically developed with constrained material; quantities of data, limiting their functionality and performance. The modest market share of AI SuSu's is perceived as weakness for enticing third-party developers. Technological complexity, in the form of micro-pixels and fake data entries, are utilised. These measures simplify auditing processes, protecting IP by threating litigation.

IP can be strategically disclosed during client or investor negotiations. It is believed to be essential to disclose pieces of IP during negotiations as clients and investors question whether the product does indeed make use of AI. Moreover, AI SuSu's must disclose pieces of IP to supplement limited understandings of potential clients and investors.

In general, formal appropriability mechanisms are deemed ineffective as, once legal protection has been established, IP infringement is deemed near impossible as AI-related IP is too ambiguous. Copyrights see disuse as the tacitness of information makes legally defining IP unfeasible. Additionally, extensive open-source assets usage conditions copyright usage because some creative commons simply prohibit legal protection of resulting products. Key resources shift towards the data pipeline makes database rights more relevant in the eyes of AI SuSu's. However, usage of these rights is deterred as the description of the conditions, needed to receive database rights, are unclear to AI SuSu's making them hesitant to depend on this instrument. AI models differ in their dependence on data, this in turn affecting propensities for utilising database rights. Trade secrecy is utilised sparsely, in the form of employee contracts, NDAs with clients, and license agreements with service providers.

Five data-driven themes emerged. Mega databases are built to allow for re-training and comparison. Additionally, products for other purposes or clients can be developed using these databases. Data storage and model training is mostly done via third-party services. This reliance undermines profitability and have possible implications for IP strategies. Some AI SuSu's overlook the intricacies of creative commons licenses for lack of manpower or because the chance of getting caught is believed insignificant. AI-related IP is strategically disclosed due to the side-effects of ambiguity. Clients or partners either lacking understanding or doubting whether AI was truly used, demand explanations of the processes leading from data to prediction. Sharing this crucial information raises risks of IP, which are addressed by strategically disclosing IP. Excessive amounts of information are provided, convincing clients or partners, while concealing key details. Market novelty and unique characteristics of AI-related IP push current legislation to its limits. As a consequence, legislation has yet reached consensus on crucial topics such effort and the eligibility of synthetic or derivate databases, causing disuse of formal mechanisms.

6. Discussion

6.1. Limitations

Notwithstanding the answer to the research question, the study contains some methodological limitations. The final sample size remained relatively small due to an initial overestimation of response rates. Furthermore, the economics effects of the COVID-19 pandemic may have caused SuSu's to go bankrupt, shrinking the available sample size. The relatively small sample size, compared to the greater population, threatens external validity. A more representative picture of IP strategies would have been realized if the sample size was increased. Also, the study was conducted for the NL AI Coalition, therefore the sample was exclusively composed of Dutch-based AI SuSu's. Including AI SuSu's from other regions would have yielded more generalised results, less dependent on region-specific factors (e.g., local legislation, culture, etc.). Applying the framework to AI SuSu's from other regions tackles this limitation, assessing the validity of themes and sharpening them accordingly.

Also included were several theoretical limitations. Since the research was exploratory, meant to gain an initial understanding of AI SuSu IP strategies, a basic analytical framework was used. As a result, the effects of themes were assessed in isolation, ignoring interactions between themes. Taking these interactions into consideration would have produced more valid findings, both internally and externally. A simplified overview of AI development was used, to ensure themes applied cross-cases, supporting internal validity. However, findings revealed two AI development aspects; data pipeline and third-party service usage, which were initially overlooked but when included in the study would have generated a more representative depiction of AI SuSu IP strategies. Subsequent research incorporating these findings might address this limitation.

6.2. Theoretical implications

Calvin et al. (2020) suggested to investigate IP strategies utilised by AI stakeholders, other than large US corporates. Our study is the first to shed light on IP strategies, in the context of AI SuSu's, by utilising theories derived from related industries. Findings revealed these theories to generally hold, suggesting this theoretical foundation to be compatible in the context of AI. Addressing the literary gap provides insights relevant to AI-specific literature and the greater body of IP strategy literature.

The descriptions of the IP strategies utilised by AI SuSu's may act as a springboard for subsequent innovation science research. Directing attention to individual appropriability mechanisms and factors may hopefully offer additional insights.

Key resources were defined as data and algorithm, hardware, and data by Wachowicz & Gonçalves (2019). However, findings extend these definitions by including data collection and model hosting. Previous factors behind partially disclosing information in software related to signalling firm value and applying for patents (Baccara & Razin, 2012). An additional motivation emerged to disclose IP, to convince potential clients and prove AI was truly utilised.

General IP strategy theories are validated, assessing whether hypotheses derived from software SuSu's remain valid in the context of AI. Past scholars studied the effects of shifting key resources on IP strategies (Blind et al., 2006; Chatterjee, 2017b; Tantleff, 2015a; M Wachowicz & Gonçalves, 2019), our study extends this line of reasoning, illustrating additional implications when moving from moving from software to AI products. Similarly, extent literature on tacitness by Hall et al. (2014) is validated since findings corroborated its negative impact on IP strategies. Theories on the effects of R&D intensities by Veugelers & Schneider (2018 cannot be confirmed nor invalidated due to contradictory results, confounded product maturity.

Lastly, general IP strategy theory is extended, providing novel insights thanks to data-driven themes. One, the concept of ambiguity, traditionally seen as a barrier preventing firms from utilising certain appropriability mechanisms (Fromer, 2019; Kearns & Lederer, 2003), deserves greater recognition in future studies on AI IP strategies. In the context of AI, ambiguity is embraced, used as a tool against infringement. Hence, recognising the effects of ambiguity becomes a prerequisite for fully understanding IP strategies in the context of AI. Furthermore, recognising these effects can be beneficial for studies on technologies with similar IP-related characteristics. For example, to study IP in quantum computing applications (Cusumano, 2018). Two, market newness posited by Chen et al. (2005) can be extended by acknowledging the novelty of legislation. This is particularly relevant for AI, as traditional conceptualisation of author and ownership are upturned (Calvin et al., 2020). Operationalising the novelty of legislation is left for future research.

6.3. Managerial implications

Next to additions to theory, the study provides important insights for managers and policy makers.

Primarily, findings lower IP-related uncertainties. Results illustrate which appropriability mechanisms are available and most conducive in AI. AI SuSu's can use these findings to design more suitable IP strategies, being certain of its effectiveness. Delponte (2018) argued that uncertainties can be reduced by understanding their sources. Results provide initial understanding of these sources by illustrating how innovation-specific and market factors influence IP strategy decision-making.

Crucial AI-related IP data and models are accessible electronically, threatening the ability for keeping information confidential. To this end, it is discouraged to provide employees with remote electronic access to these assets (Hemphill, 2004). This becoming especially relevant considering recent "working from home" trends (Ozimek, 2020). Managers do well by understanding the importance of data and models, ensuring only sections, relevant to core tasks, are accessible by employees. Findings by Klein (2020) who found software SuSu's to utilise ambiguity as a tool against infringement are corroborated by our results. Managers benefit by better understanding this relationship, appreciating its usefulness against infringement. If permitted by use-case and budget, managers may strengthen ambiguity by delivering AI products via the cloud or edge-devices. Reverse engineering and IP theft is dissuaded as other firms lack direct access to models and data. Conversely, when use-cases or budget make this approach infeasible, managers can dimmish ambiguity, adopting “explainable” AI model varieties. These models more clearly illustrate how data led to a prediction, reducing ambiguity. When done sufficiently, would allow AI SuSu’s to apply for a computer implemented invention (EPO, 2021).

AI SuSu’s can explore strategies to become less reliant on third parties, defining how the firm aims to procure its own hosting and machine learning hardware. Alternatively, the AI NL Coalition might explore offering these types of services. This tackles two issues. One, assuring AI SuSu’s stable fees diminished uncertainties about potential price increases threatening profitability. Two, many potential clients were shown to demand data to be hosted and processed within national borders. This demand can currently not always be fulfilled because some third-party services cannot guarantee data remains within European confines. Providing domestic and inexpensive alternatives to third-party services addresses these issues, in turn providing AI SuSu’s access to larger pools of potential clients.

Findings have direct implications for database rights policies. This mechanism is considered relevant by AI SuSu’s, but cases revealed that unclear descriptions of the necessary conditions stifle its application. Policymakers must clarify the definition of effort, needed to receive database rights, taking several AI-specific topics into account. For example, the eligibility of derivative datasets, generated by combining two or more public databases. Or under which conditions synthetic or AI-generated databases can be protected using database rights.

In sum, findings suggest secrecy, trade-secrecy, and lead-times are currently most conducive to AI SuSu’s. Subsequent researchers might use these findings as a springboard to individually examine ways these mechanisms can be applied by AI SuSu’s. Firstly, by studying the implications of tacitness, ambiguity, and open-source asset usage on strategic secrecy management. Secondly, by considering under which conditions AI-related IP is best protect via trade-secrecy. Utilising longitudinal case studies will provide data on specific causal relationships between these factors, gathering data from business documents, interviews, and observations.

Furthermore, additional research studying the implications of data-driven themes might yield more representative depictions of AI SuSu IP strategies. For instance, by including data collection and model distribution as key resources, accounting for the effects of third-party service reliance on IP strategies and investigating when AI-related ambiguity is sufficiently reduced for CII applications.

7. References

Amazon. (2021). *Amazon's AutoGluon helps developers deploy deep learning models with just a few lines of code*. https://www.amazon.science/amazons-autogluon-helps-developers-get-up-and-running-with-state-of-the-art-deep-learning-models-with-just-a-few-lines-of-code

Anderson, K., Burford, O., & Emmerton, L. (2016). Mobile health apps to facilitate self-care: A qualitative study of user experiences. *PLoS ONE*, *11*(5), 1–22. https://doi.org/10.1371/journal.pone.0156164

Ayoub, K., & Payne, K. (2016). Strategy in the Age of Artificial Intelligence. *Journal of Strategic Studies*, *39*(5–6), 793–819. https://doi.org/10.1080/01402390.2015.1088838

Baccara, M., & Razin, R. (2012). Bargaining Over New Ideas: Rent Distribution and Stability of Innovative Firms. *SSRN Electronic Journal*, *5*(6), 1095–1129. https://doi.org/10.2139/ssrn.1124944

BG.legal. (2021). *BG.legal*.

Blind, K., Edler, J., Frietsch, R., & Schmoch, U. (2006). Motives to patent : Empirical evidence from Germany. *Research Policy*, *35*(3), 655–672. https://doi.org/10.1016/j.respol.2006.03.002

Bloomberg. (2020). BrainBox AI Recognized by TIME as a Best Invention of 2020 [Press release]. In *Bloomberg Business*. https://www.bloomberg.com/press-releases/2020-11-19/brainbox-ai-recognized-by-time-as-a-best-invention-of-2020

Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage.

Bradley, S. W., Kim, P. H., Klein, P. G., McMullen, J. S., & Wennberg, K. (2021). Policy for innovative entrepreneurship: Institutions, interventions, and societal challenges. *Strategic Entrepreneurship Journal*, *15*(2), 167–184. https://doi.org/10.1002/sej.1395

Braun, V., & Clarke, V. (2012). Thematic analysis. *APA Handbook of Research Methods in Psychology, Vol 2: Research Designs: Quantitative, Qualitative, Neuropsychological, and Biological.*, *2*, 57–71. https://doi.org/10.1037/13620-004

Bryman, A. (2016). *Social research methods*. Oxford University Press.

Calvin, N., Leung, J., Anderljung, M., Brundage, M., Dafoe, A., Ding, J., Flynn, C., Hadshar, R., Ism, C., Khan, S., Tse, B., Zhang, B., & Zwetsloot, R. (2020). *Who owns artificial intelligence? A preliminary analysis of corporate intellectual property strategies and why they matter Summary and Key Takeaways* (Issue February).

Canfora, G., & Di Penta, M. (2007). New frontiers of reverse engineering. *FoSE 2007: Future of Software Engineering*, 326–341. https://doi.org/10.1109/FOSE.2007.15

Chaiklin, H. (1991). Case Study Research: Design and Methods. *The Journal of Nervous and Mental Disease*, *179*(2), 115. https://doi.org/10.1097/00005053-199102000-00025

Chatterjee, J. (2017a). Strategy, human capital investments, business-domain capabilities, and performance: a study in the global software services industry. *Strategic Management Journal*, *38*(3), 588–608. https://doi.org/10.1002/smj.2505

Chatterjee, J. (2017b). Strategy, human capital investments, business-domain capabilities, and performance: A study in the global software services industry. *Strategic Management Journal*, *38*, 588–608.

Chen, J., Reilly, R. R., & Lynn, G. S. (2005). The impacts of speed-to-market on new product success: The moderating effects of uncertainty. *IEEE Transactions on Engineering Management*, *52*(2), 199–212. https://doi.org/10.1109/TEM.2005.844926

Cockburn, I. M., & MacGarvie, M. J. (2011). Entry and patenting in the software industry.

Management Science, *57*(5), 915–933. https://doi.org/10.1287/mnsc.1110.1321
Cohen, W. M., Nelson, R. R., & Walsh, J. P. (2000). Protecting Their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not). *NBER Working Papers*, *7552*, 50. https://doi.org/10.3386/w7552
Creative Commons. (2021). *About The Licenses*. https://creativecommons.org/licenses/?lang=en
Cusumano, M. A. (2018). Technology strategy and management: The business of quantum computing. *Communications of the ACM*, *61*(10), 20–22. https://doi.org/10.1145/3267352
de Laat, P. B. (2005). Copyright or copyleft ? An analysis of property regimes for software development. *Research Policy*, *34*, 1511–1532. https://doi.org/10.1016/j.respol.2005.07.003
Delponte, L. (2018). European Artificial Intelligence (AI) leadership, the path for an integrated vision. *European Parliament*, *September*, 48. http://www.europarl.europa.eu/RegData/etudes/STUD/2018/626074/IPOL_STU(2018)626074_EN.pdf
Dubois, A., & Gadde, L. E. (2002). Systematic combining: An abductive approach to case research. *Journal of Business Research*, *55*(7), 553–560. https://doi.org/10.1016/S0148-2963(00)00195-8
Energy Central. (2020). *The Rise of Artificial Intelligence in Renewable Energy - 2020*. https://energycentral.com/c/cp/rise-artificial-intelligence-renewable-energy-2020#:~:text=The field of renewable energy,accurate weather and load forecasting.
EPO. (2021). *Index for Computer-Implemented Inventions*. https://www.epo.org/law-practice/legal-texts/html/guidelines/e/j.htm
Flewitt, R., Messer, D., & Kucirkova, N. (2015). New directions for early literacy in a digital age: The iPad. *Journal of Early Childhood Literacy*, *15*(3), 289–310. https://doi.org/10.1177/1468798414533560
Fromer, J. C. (2019). Machines as the new oompa-loompas: Trade secrecy, the cloud, machine learning, and automation. *New York University Law Review*, *94*(4), 706–736.
Giardino, C., Unterkalmsteiner, M., Paternoster, N., Gorschek, T., & Abrahamsson, P. (2014). What do we know about software development in startups? *IEEE Software*, *31*(5), 28–32. https://doi.org/10.1109/MS.2014.129
Graham, S. J. H., & Sichelman, T. S. (2016). Intellectual property and technology startups: What entrepreneurs tell us. In *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth* (Vol. 26, pp. 163–199). Emerald Group Publishing Limited. https://doi.org/10.1108/S1048-473620160000026006
Grand View Research. (2020). *Artificial Intelligence Market Size, Share & Trends Analysis Report By Solution (Hardware, Software, Services), By Technology (Deep Learning, Machine Learning), By End Use, By Region, And Segment Forecasts, 2020 - 2027*. https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market
Guttentag, D. (2015). Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector. *Current Issues in Tourism*, *18*(12), 1192–1217. https://doi.org/10.1080/13683500.2013.827159
Gyimóthy, T., Ferenc, R., & Siket, I. (2005). Empirical validation of object-oriented metrics on open source software for fault prediction. *IEEE Transactions on Software Engineering*, *31*(10), 897–910. https://doi.org/10.1109/TSE.2005.112
Hall, B., Helmers, C., Rogers, M., & Sena, V. (2014). The Choice Between Formal and Informal

Intellectual Property: A Literature Review. *Journal of Economic Literature*, *52*(2), 375–423. https://doi.org/10.1257/jel.52.2.375

Hall, B., & Ziedonis, R. (2001). The Patent Paradox Revisited: An Empirical Study of Patenting in the US Semiconductor Industry, 1979-95. *RAND Journal of Economics*, *32*(1), 101–128. https://doi.org/10.2307/2696400

Harrison, S., & Sullivan, P. (2006). *Einstein in the Boardroom*. John Wiley & Sons, Inc.

Hein, A., Schreieck, M., Riasanow, T., Setzke, D. S., Wiesche, M., Böhm, M., & Krcmar, H. (2020). Digital platform ecosystems. *Electronic Markets*, *30*(1), 87–98. https://doi.org/10.1007/s12525-019-00377-4

Heirman, A., & Clarysse, B. (2004). The initial resources and market strategy to create high growth firms. *Working Paper Steunpunt OOI*, *October*, 1–64.

Hemphill, T. (2004). The Strategic Management of Trade Secrets in Technology-based Firms. *Technology, Analysis & Strategic Management*, *16*(4), 479–494. https://doi.org/10.1080/0953732042000295793

Henry, E., & Ruiz-Aliseda, F. (2012). Innovation Beyond Patents: Technological Complexity as a Protection against Imitation. *CEPR Discussion Paper*, *8870*, 1–41. https://ssrn.com/abstract=2034088

Hilmola, O. P., Helo, P., & Ojala, L. (2003). The value of product development lead time in software startup. *System Dynamics Review*, *19*(1), 75–82. https://doi.org/10.1002/sdr.255

Hilty, R. M., Hoffman, J., & Scheuerer, S. (2020). Intellectual Property Justification for Artificial Intelligence. *Max Planck Institute for Innovation and Competition Research Paper*, *20*(02), 1–29. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3539406

Holloway, I., & Todres, L. (2003). The Status of Method: Flexibility, Consistency and Coherence. *Qualitative Research*, *3*(3), 345–357. https://doi.org/10.1177/1468794103033004

James, S., Leiblein, M., & Lu, S. (2013a). How Firms Capture Value From Their Innovations. *Journal of Management*, *39*(5), 1123–1155.

James, S., Leiblein, M., & Lu, S. (2013b). How Firms Capture Value From Their Innovations. *Journal of Management*, *39*(5), 1123–1155. https://doi.org/10.5465/ambpp.2013.17583abstract

Kearns, G. S., & Lederer, A. L. (2003). A resource-based view of strategic IT alignment: How knowledge sharing creates competitive advantage. *Decision Sciences*, *34*(1), 1–29. https://doi.org/10.1111/1540-5915.02289

Kerravala, Z. (2019). *Why Wi-Fi needs artificial intelligence*. Network Intelligence. https://www.networkworld.com/article/3355237/why-wi-fi-needs-artificial-intelligence.html

Kessler, E. H., & Chakrabarti, A. K. (1996). Innovation Speed : A Conceptual Model of Context, Antecedents, and Outcomes. *The Academy of Management Review*, *21*(4), 1143–1191. https://www.jstor.org/stable/259167

Kim, M. (2007). The creative commons and copyright protection in the digital era: Uses of creative commons licenses. *Journal of Computer-Mediated Communication*, *13*(1), 187–209. https://doi.org/10.1111/j.1083-6101.2007.00392.x

Klein, M. A. (2020a). Secrecy, the patent puzzle and endogenous growth. *European Economic Review*, *126*. https://doi.org/10.1016/j.euroecorev.2020.103445

Klein, M. A. (2020b). Secrecy, the patent puzzle and endogenous growth. *European Economic Review*, *126*, 103445. https://doi.org/10.1016/j.euroecorev.2020.103445

Kochar, M. (2020). Open Source AI and Data: How best to keep up with rapid advances? *IBM Developer*. https://developer.ibm.com/blogs/open-source-ai-and-data-how-best-to-keep-up-with-rapid-advances/
Kop, M. (2020a). Machine Learning & EU Data Sharing Practices. *Transatlantic Antitrust and IPR Developments*, *1/2020*.
Kop, M. (2020b). Machine Learning and EU Data-Sharing Practices: Legal Aspects of Machine Learning Training Datasets for AI Systems. In *Research Handbook on Big Data Law* (pp. 431–452). Edward Elgar Publishing.
Kulkarni, R. H., & Padmanabham, P. (2017). Integration of artificial intelligence activities in software development processes and measuring effectiveness of integration. *IET Software*, *11*(1), 18–26. https://doi.org/10.1049/iet-sen.2016.0095
LaRoque, S. J. (2018). Reverse Engineering and Trade Secrets in the Post-Alice World. *Kansas Law Review*, *66*. https://doi.org/10.17161/1808.25704
Laurell, C., & Sandström, C. (2016). Analysing uber in social media - Disruptive technology or institutional disruption? *International Journal of Innovation Management*, *20*(5). https://doi.org/10.1142/S1363919616400132
Law, K. K. (2014). The problem with knowledge ambiguity. *European Management Journal*, *32*(3), 444–450. https://doi.org/10.1016/j.emj.2013.07.010
Lerner, J., & Tirole, J. (2016). Some Simple Economics of Open Source. *The Journal of Industrial Economics*, *50*(2), 197–234. https://doi.org/10.1111/1467-6451.00174
Levin, R., Klevorick, A., Nelson, R., Winter, S., Gilbert, R., & Griliches, Z. (1987). Appropriating the Returns from Industrial Research and Development. *Competition Policy International*, *9*(2), 160–196.
Levine, D. S., & Sichelman, T. (2018). Why do startups use trade secrets? *Notre Dame Law Review*, *94*(2), 751–820. https://doi.org/10.2139/ssrn.3166834
Lins, S., Pandl, K. D., Teigeler, H., Thiebes, S., Bayer, C., & Sunyaev, A. (2021). Artificial Intelligence as a Service: Classification and Research Directions. *Business and Information Systems Engineering*, 1–16. https://doi.org/10.1007/s12599-021-00708-w
Lynn, G. S., & Akgün, A. E. (1998). Innovation strategies under uncertainty: a contingency approach for new product development. *Engineering Management Journal*, *10*(3), 11–18. https://doi.org/10.1080/10429247.1998.11414991
Mattioli, M. (2014). Disclosing big data. *Minnesota Law Review*, *99*(2), 535–584.
Maurer, S. M., Hugenholtz, P. B., & Onsrud, H. J. (2001). *Europe's database experiment*. 789–790. https://doi.org/10.1126/science.1062695
McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. C., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F. J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C. J., King, D., … Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, *577*(7788), 89–94. https://doi.org/10.1038/s41586-019-1799-6
Microsoft. (2021). *AI platform, tools, and services*. https://www.microsoft.com/en-us/ai/ai-platform
Miller Cole, B. (2019). *When Does A Start-Up Become A Scale-Up?* Forbes.
Mockus, A., & Herbsleb, J. D. (2000). Two Case Studies of Open Source Software Development: Apache and Mozilla. *ACM Transactions on Software Engineering and Methodology*, *11*(3), 263–272. https://doi.org/10.1145/567793.567795
Muehlhauser, L., & Riber, L. (2014). Exponential and non-exponential trends in information technology. In *Machine Intelligence Research Institute*.

NL AIC. (2021). *Dutch AI Coalition*. https://nlaic.com/en/
Nonaka, I., & von Krogh, G. (2009). Tacit knowledge and knowledge conversion: Controversy and advancement in organizational knowledge creation theory. *Organization Science*, *20*(3), 635–652. https://doi.org/10.1287/orsc.1080.0412
Ozimek, A. (2020). *Economist Report: Future Workforce [Press release]*. https://www.upwork.com/press/releases/economist-report-future-workforce
Paternoster, N., Giardino, C., Unterkalmsteiner, M., Gorschek, T., & Abrahamsson, P. (2014). Software development in startup companies: A systematic mapping study. *Information and Software Technology*, *56*(10), 1200–1218. https://doi.org/10.1016/j.infsof.2014.04.014
Pike, C. G. (2001). *Virtual Monopoly: Building an Intellectual Property Strategy for Creative Advantage-from Patents to Trademarks, from Copyrights to Design Rights*. Nicholas Brealey Publishing.
Pisano, G., & Teece, D. (2007). How to Capture Value from Innovation: Shaping Intellectual Property and Industry Architecture. *California Management Review*, *50*(1), 278–296.
Pujol, N. (2012). Freemium: Attributes of an Emerging Business Model. *SSRN Electronic Journal*, *December*. https://doi.org/10.2139/ssrn.1718663
QSR International Pty Ltd. (2020). *NVivo*.
Rijksoverheid. (2020). *Nederlandse Digitaliseringsstrategie 2.0*. https://www.rijksoverheid.nl/documenten/rapporten/2019/07/05/nederlandse-digitaliseringsstrategie-2.0
Samuelson, P. (2010). Legally speaking: Why do software startups patent (or not)? *Communications of the ACM*, *53*(11), 30–32. https://doi.org/10.1145/1839676.1839687
Samuelson, P., & Scotchmer, S. (2002). The Law and Economics of Reverse Engineering. *The Yale Law Journal*, *111*(7), 1575–1663. https://www.jstor.org/stable/797533
Stancin, I., & Jovic, A. (2019). An overview and comparison of free Python libraries for data mining and big data analysis. *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 977–982. https://doi.org/10.23919/MIPRO.2019.8757088
Svoboda, E. (2020, November). Artificial intelligence is improving the detection of lung cancer. *Nature*. https://www.nature.com/articles/d41586-020-03157-9
Taipale, M. (2010). Huitale - a story of a Finnish lean startup. *International Conference on Lean Enterprise Sotware and Systems*, 111–114. https://doi.org/10.1007/978-3-642-16416-3_16
Tantleff, A. (2015a). Considerations on big data licensing. *Managing Intellectual Property*, *246*(14), 2015.
Tantleff, A. (2015b). Considerations on big data licensing. *Managing Intellectual Property*, *246*(14), 2015. https://heinonline.org/HOL/P?h=hein.journals/manintpr246&i=16
TechLeap. (2020). *Startups & scaleups*. https://finder.techleap.nl/companies.startups/f/all_locations/allof_Netherlands/locations/anyof_Netherlands/technologies/anyof_artificial intelligence
Teece, D. J. (2018a). Profiting from innovation in the digital economy: Enabling technologies, T standards, and licensing models in the wireless world. *Research Policy*, *47*(8), 1367–1387. https://doi.org/10.1016/j.respol.2017.01.015
Teece, D. J. (2018b). Profiting from innovation in the digital economy: Enabling technologies, T standards, and licensing models in the wireless world. *Research Policy*, *47*(8), 1367–1387.

Timmermans, S., & Tavory, I. (2012). Theory construction in qualitative research: From grounded theory to abductive analysis. *Sociological Theory*, *30*(3), 167–186. https://doi.org/10.1177/0735275112457914

TNO. (2020). *Effective interventions to increase energy efficiency and reduce energy poverty*. https://www.tno.nl/en/focus-areas/energy-transition/roadmaps/towards-broad-support-for-the-energy-transition/the-social-aspects-of-the-energy-transition/effective-interventions-to-increase-energy-efficiency-and-reduce-energy-poverty/

Vasilev, I., Slater, D., Spacagna, G., Roelants, P., & Zocca, V. (2019). *Python Deep Learning: Exploring deep learning techniques and neural network architectures with PyTorch, Keras, and TensorFlow*. Packt Publishing Ltd.

Veugelers, R., & Schneider, C. (2018). Which IP strategies do young highly innovative firms choose? *Small Business Economics*, *50*(1), 113–129. https://doi.org/10.1007/s11187-017-9898-y

von Hippel, E. (2005). Open Source Software Projects as User Innovation Networks. *Perspectives on Free and Open Source Software*, 267–278. https://doi.org/10.7551/mitpress/5326.003.0021

Von Hippel, E. (2005). Open Source Software Projects as User Innovation Networks. *Perspectives on Free and Open Source Software*, 267–278. https://doi.org/10.7551/mitpress/5326.003.0021

Von Hippel, E., & Von Krogh, G. (2003). Open source software and the “private-collective” innovation model: Issues for organization science. *Organization Science*, *14*(2). https://doi.org/10.1287/orsc.14.2.209.14992

Von Krogh, G., Spaeth, S., & Lakhani, K. R. (2003). Community, joining, and specialization in open source software innovation: A case study. *Research Policy*, *32*(7), 1217–1241. https://doi.org/10.1016/S0048-7333(03)00050-7

Wachowicz, M, & Gonçalves, L. R. (2019). *Artificial Intelligence and Creativity - New Concepts in Intellectual Property*. Gedai Publicaçös.

Wachowicz, Marcos, & Ruthes Gonçalves, L. (2019). Artificial Intelligence and Creativity: New concepts in intellectual property. In *MATEC Web of Conferences*. Gedai Publicaçös. https://www.matec-conferences.org/10.1051/matecconf/201817601043

Wagner, D. N. (2020). The nature of the Artificially Intelligent Firm - An economic investigation into changes that AI brings to the firm. *Telecommunications Policy*, *44*(6), 101954. https://doi.org/10.1016/j.telpol.2020.101954

WTO. (1995). *Overview: the TRIPS Agreement*. https://www.wto.org/english/tratop_e/trips_e/intel2_e.htm#relatedright

8. Appendices

Appendix A. – Initial code manual including codes, theme definitions, and descriptions of theme presence.

Code	Label	Definition	Description
KS	Key resource shift	Compared to software SuSu's, AI SuSu's assess data as being more critical to their business.	Participants mentions how data is their more vital asset.
KS1	Copyright complexities	No legal author for written code, therefore, no indicated use or importance of copyright.	AI SuSu's make no use of copyrights as vital IP-assets are not captured under traditional copyright legislation.
KS2	Database right dominance	AI SuSu's forego copyright protection for data base rights as key resources shift from source code to processed data.	AI SuSu's indicate the usage of database rights as this is their most prized resource.
IT	Increased tacitness	The knowledge and skills used for AI development more difficult to put into words and dissemination than those used for software development.	Participants remarks how the knowledge required during development is based on experiences, skills and hard to disseminate.
IT1	Legal definition difficulties	IP in the form of experience and skills more difficult to capture in legal terms.	AI SuSu's indicate how the legal difficulties led to a reduced reliance on formal mechanisms for IP protection.
IT2	Technologically complex tacitness	Increased tacitness of knowledge makes technological complexity less suitable.	AI SuSu's confirm how the increased tacit nature of knowledge dissuades the use of technological complexity.
OS	Use of open-source assets	AI SuSu's make more pervasive usage of open-source assets compared to software SuSu's.	Subject mentions extensive use of open-source assets when discussing the product's development process.
OS1	Creative commons prohibition	Creative commons licencing forbids AI SuSu's from using legal IP protection instruments.	AI SuSu's attribute their disuse of formal mechanisms to their statistical models' creative commons licenses.
OS2	Conflicting interests	Open-source motivations clash motives for keeping IP private.	AI SuSu's assess free sharing and communal development and therefore refrain from keeping IP secret.
OS3	Open-source complementaries	As AI SuSu's increase their utilisation of open-source assets, so does their interest in third-party developed complementary assets.	The AI SuSu mentions how pervasive usage of open-source assets lead to the pursual of complementary assets.
OS4	Easy open-source reverse engineering	Pervasive open-source asset usage makes technological complexity less enticing as competitors have access the similar input assets.	AI SuSu's indicate a reluctance in making input components technologically complex due to them being publicly available to competitors.

IA	Increased ambiguity	The connection between input and output components is, at the least vague (compared to software), and at the most absent for AI products.	Participants remarks on the difficulties in tracing input and output components as these ties are often missing.
IA1	Ambiguous secrecy	Higher propensities for secrecy as AI SuSu's trust protection from ambiguity.	AI mention to utilise primarily on secrecy because competitors cannot determine the innerworkings of their product.
IA2	Novel obfuscation technologies	Expectations for novel applications of technological complexity.	AI SuSu's bring up AI-specific obfuscation technologies.
MN	Market newness	The AI market is inherently newer to software, thus AI SuSu's experience more rapid changes in consumer type and/or distribution channels.	Participant suggests that the novelty of the AI industry influenced/influences the chosen IP strategy.
MN1	Novel consumers	AI SuSu's continuously sell product to novel consumer types.	AI SuSu's indicate the need to invest in lead-times in order to keep up with quickly changing user preferences.
MN2	Novel distribution	AI SuSu's indicate the need to rapidly deliver products via previously unused distributions channels.	SuSu's mention lead-times are most important in keeping up with quickly changing channels of distribution.
RD	R&D intensity	AI's higher R&D intensity causes shifts in the type and manner of protecting IP.	AI SuSu's mentions how R&D is a major investment and suggests its influence on the chosen IP strategy.
RD1	Complex confidentiality	Higher R&D intensity, and subsequently complexity of technology, causes firms to rely more heavily on secrecy.	Firm confirms to spend more 73 per cent on R&D, and as a result opted for secrecy to protect the resultant IP.

Appendix B. - Overview AI SuSu's in sample

		Market tenure range in years					
		0 and 2	3 and 5	6 and 8	9 and 11	11+	Total
	Per market tenure	29	32	13	4	15	93
Sub-field	Data generation	1	0	0	0	0	1
	Built environment	0	1	0	0	0	1
	Business support	8	9	1	3	1	22
	Career platform	2	1	0	0	0	3
	Consultancy	0	1	1	0	2	4
	Customer service	3	1	0	0	0	4
	Cybersecurity	1	0	0	0	0	1
	Data management	0	1	0	0	0	1
	Decision-making	1	0	0	0	0	1
	Ecommerce	1	2	0	0	0	3
	Energy	1	1	0	0	0	2
	Financial services	0	0	1	0	0	1
	Food tech	1	0	0	0	0	1
	General developer	4	10	6	0	8	28
	Health	2	2	3	0	3	10
	Industry	0	1	0	0	0	1
	Predictive maintenance	0	1	0	0	0	1
	Logistics	3	0	1	1	1	6
	Science & research	1	1	0	0	0	2