



The dark side of AI in professional services

Francisco J. Trincado-Munoz, Carlo Cordasco & Tim Vorley

To cite this article: Francisco J. Trincado-Munoz, Carlo Cordasco & Tim Vorley (10 Apr 2024): The dark side of AI in professional services, The Service Industries Journal, DOI: [10.1080/02642069.2024.2336208](https://doi.org/10.1080/02642069.2024.2336208)

To link to this article: <https://doi.org/10.1080/02642069.2024.2336208>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 10 Apr 2024.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

The dark side of AI in professional services

人工智能在专业服务领域的阴暗面

Francisco J. Trincado-Munoz^a, Carlo Cordasco^{a,b} and Tim Vorley^a

^aOxford Brookes Business School, Oxford Brookes University, Oxford, UK; ^bAlliance Manchester Business School, University of Manchester, Manchester, UK

ABSTRACT

The introduction and widespread adoption of Artificial Intelligence in the professions has the potential to deliver a number of critical public goods, such as widening access to justice and healthcare through AI-powered professional services. Yet, the deployment of AI in the professions does not come without challenges, exemplified by the concerns about explainability, privacy, and human agency. In this paper, we explore how these issues may give rise to dark sides of AI in professional services and illustrate how an uncoordinated process of adoption and deployment can threaten the scope of AI-powered services. In particular, we illustrate how the adoption and deployment of AI in services may undermine the fiduciary duty between clients and professionals that, so far, has safeguarded the relationship between them, creating a 'market for lemons' of professional services. We conclude with a reflection on plausible ways forward to facilitate and smooth the transition to AI-powered services.

摘要

人工智能在专业服务领域的引入及广泛应用具有为公众提供一些重要服务的潜力。例如通过人工智能驱动的专业服务扩大对公正和健康的使用所能达到的范围。然而，人工智能在专业服务领域的应用并非没有挑战，例如对可解释性，隐私及人类施为的担忧。在本文中，我们探讨这些问题可能如何引发人工智能在专业服务领域应用中的阴暗面的挑战，并且展示一个未经协调的采用及应用过程如何能够威胁人工智能驱动的服务的规模。特别需要指出的是，我们展示了采纳和应用人工智能会如何破坏客户和专业服务人员之间的信托责任，形成一种“柠檬市场次品市场”。这个信托责任到目前为止确保了他们客户和专业服务人员之间的关系。我们在结论中显示了促成及使之过渡到人工智能驱动服务顺畅的可能途径。

ARTICLE HISTORY

Received 14 September 2023
Accepted 25 March 2024

KEYWORDS

AI; dark sides; professions;
professional services;
adoption of AI; LLMs

关键词

人工智能; 专业; 专业服务;
人工智能的的采用; LLMs

CONTACT Francisco J. Trincado-Munoz  ftrincado-munoz@brookes.ac.uk  Oxford Brookes Business School, Oxford Brookes University, Headington Campus, Headington Rd, Headington, Oxford OX3 0BP, UK

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Introduction

The promise of Artificial Intelligence (AI) in Professional Services is one to look forward to, with faster, more efficient, and more accurate professional judgment potentially widening access to crucial public goods such as justice and healthcare (Acemoglu & Restrepo, 2018; Susskind & Susskind, 2015). Professional Service Firms (PSFs) are knowledge-intensive organisations that master a particular expertise or knowledge base (Von Nordenflycht, 2010) through highly skilled, expert workers (Spring et al., 2022). Commonly, financial, legal, accountancy, management consultancy, and advertising firms are considered PSF (Trincado-Munoz et al., 2023), yet the scope of professional firms can be even broader (Von Nordenflycht, 2010). AI adoption in Professional Services does not come without challenges (Belanche et al., 2020; R. W. Belk et al., 2023; Brooks et al., 2020), partly stemming from the current lack of technological maturity of AI technologies, alongside an increasing resistance among professionals, who are accustomed to certain sets of practices and routines, and who also enjoy the responsibilities and rewards attached to their professional identity (Cordasco et al., 2021; Faulconbridge et al., 2023; Perner & Werr, 2023; Spring et al., 2022).

As AI is becoming more pervasive and relevant, it needs to become more reliable. Misuse, and overuse, of AI might carry significant risks for users and the community, which prompts important concerns and the necessity to be cautious. For this reason, AI adoption in Professional Services has exacerbated the debates about explainability, privacy, and agency (Belk, 2021; Jin, 2018; Park et al., 2021). These three aspects can potentially disrupt the provision of services, affecting organisations, professionals, end-users, and society as a whole. In practice, they can create challenges such as reduced access to services, marginalisation and discrimination, malicious use of AI, detrimental quality of services, data security breaches, and inaccuracy of services.

The development and implementation of AI in professional services are not exempt from ethical considerations. Unethical AI could breed distrust, resentment, and unnecessary conflict (Véliz, 2019). Furthermore, AI has the potential to augment dark practices in management, blurring the lines between what is right and what is wrong (Belanche et al., 2024; Wagner, 2021). AI tools are powerful enough that their misuse can cause great harm and disrupt the way the professional practice functions. In this article, we explore the potential dark sides of AI adoption in professional services. We also illustrate how an uncoordinated adoption process could lead to a state of affairs in which AI, far from delivering societal benefits, gives rise to a 'market for lemons' (Akerlof, 1970) in the services sectors. In doing so, we look at professional services from an institutional standpoint (North, 1991; Scott, 1994), in an effort to highlight how formal and informal rules, underpinning professional identity, and fiduciary duty, have, to now, safeguarded the trust between clients and professionals by solving the principal-agent problem generated by informational asymmetries (Brien, 1998; Easterbrook & Fischel, 1993; Sealy, 1962).

AI adoption in Professional Services, along with the emergence of readily available AI tools providing pseudo-professional advice (e.g. Large Language Models (LLMs) such as ChatGPT), may pose a bigger challenge to the trusting relationship between clients and professionals, one that may undermine, if not eliminate, the potential benefits of AI-powered services. For instance, consider the recent case of two New York lawyers fined for having submitted a legal brief that included six fictitious case citations generated by ChatGPT (Merken, 2023). How is the public going to react to an uncoordinated process

of AI adoption that might expose a lack of professionalism? This may not merely undermine the trust between clients and professionals but also elicit, or reinforce, the belief that there is no significant difference between AI-generated and professional advice.

Despite multiple studies addressing the potential negative consequences of AI and new technologies (Acemoglu, 2021; Belanche et al., 2024; Flavián et al., 2024; Floridi et al., 2018; Susskind & Susskind, 2015; Véliz, 2021a), not much has been said with respect to their concrete and specific effects on professional services. In the same way, a narrative able to spell out how the emergence of these threats might unfold is notably missing. While existing studies tend to focus on the upsides of technology adoption in tasks and processes (see Leitner-Hanetseder et al., 2021; Perner & Werr, 2023; Spring et al., 2022; Tredinnick, 2017), our contribution lies in illustrating how an uncoordinated process of adoption within professions may betray the many promises of AI and, consequently, lead us toward a sub-optimal equilibrium.

The goal of this paper is thus two-fold. First, we illustrate how AI challenges apply to the professional services domain and may give rise to dark sides that have the potential to affect a wide array of stakeholders as well as generate negative consequences for society as a whole. Although we make no pretence of providing novel insights or depth of analysis in illustrating the dark sides of AI adoption, it is important to summarise and relate them to the professional services domain in order to form a comprehensive picture of the kind of issues the professions may be facing in the coming years. Secondly, we provide a novel narrative of how such dark sides might come to materialise, one that rests on the erosion of trust between professionals and clients that has thus far safeguarded the quality of service delivery.

This paper is organised into four main sections. In the first, we explore how foundational issues linked to AI and machine learning give rise to dark sides in professional services; in the second, we look at how professional services have traditionally dealt with informational asymmetries that create principal-agent problems between clients and professionals; in the third section, we illustrate how AI posits an even more significant challenge to the trusting relationship between clients and professionals, which may lead to a market for lemons for professional services; in the fourth section, we reflect on the way forward to favour a smooth transition to AI-powered services.

Exploring the issues behind the dark sides of AI

The exponential growth of AI and, in particular, machine learning techniques, which are the focus of this article (Belk et al., 2023; Trappenberg, 2019), carries significant uncertainties that make it urgent to think about unanticipated risks. Hence, identifying potential ethical issues, or dark sides, of AI is essential as the technology's implications may become more evident after long-term use. In this respect, we contend that the dark sides of AI are the culmination of intended and unintended, sub-optimal, or even adverse, outcomes for professionals, clients, and society. The root of these dark sides can be traced to three main issues: bias and explainability, privacy and governance, and agency.

Biases and explainability: the 'black box' problem

A significant challenge in AI systems and technology is related to the limited understanding of the process that translates inputs to outputs. As AI systems become able to process

and analyse more data, their outputs become more difficult to interpret and explain. This is also known as the 'black box' problem (Floridi et al., 2018). While the black box is a threat for everyone, it is especially salient for users who do not have access to the inner workings of the technology, creating a considerable risk associated with the use of the algorithms or the ability to understand AI-enabled decisions.

Many consequences have been linked to the black box problem. Lacking the understanding of the algorithms and/or the data used to train these can lead to marginalisation, discrimination, and unfair outcomes for end-users (Véliz, 2021a, 2021b). For instance, the uncertainty in the functioning and design of the systems can lead to disruptions in the accuracy of specific tasks, the types and weights of the criteria applied in the decision-making, or compliance with the regulation. Furthermore, in professional services, the black-box problem can affect whether or not services meet customers' expectations, present vulnerabilities, or are robust enough when they recommend a course of action (e.g. outputs are verified).

Privacy and governance

As many have highlighted, a significant threat of AI technologies is data protection, privacy, and security (Jin, 2018; Park et al., 2021; Véliz, 2021b). Since data is at the centre of AI (in development, testing, and implementation), it can create a substantial risk to companies' and users' privacy and security. For instance, in areas such as healthcare and finance, where AI systems may use personal data to produce recommendations, sensitive insights about individuals can be disclosed in case of a system/data breach, potentially exposing vulnerable groups to harm.

Further, these concerns are related to issues of data governance and accountability; on the one hand, the use of unauthorised data to produce or test technology by companies, and, on the other hand, the power held by consumers to decide how and when technology providers use their data and profit from it. Belk (2021) explains that people may not realise how many AI-powered services they already use, a situation that, to an extreme, can jeopardise human rights, dignity, and safety. The presence of AI is sometimes imperceptible to consumers (Belanche et al., 2024), creating opportunities for AI to be used wrongly, harming people, firms, and governments.

Indeed, many organisations are reluctant to disclose the data they use for business reasons or privacy concerns. Such approaches can ultimately affect the validity of the systems, creating distrust in the public and questioning the robustness of the decisions made by AI-powered systems. Belanche et al. (2024) discusses the example of the insurance industry, where the complexity of the algorithms used to calculate consumer premiums makes impossible to explain resulting decisions. In fact, the lack of clarity and understanding of the functioning and data sources also affects the ability to ensure accountability when this is needed.

Agency

Kaplan and Haenlein (2020) summarise this issue well:

Who is to blame if an AI system makes a decision that generates some harm? The magician who develop the underlying algorithm? The manufacturer who produced the software? The database that provided the external data for the system to learn from, or the customer who purchased and used the system? (p. 44)

One of the most significant concerns in AI is the lack of accountability and agency of the actors involved in the use of AI technology. This considerable issue can lead to multiple problems, some rooted in distortions generated by a conflict between principal and agents (Eisenhardt, 1989). For instance, from the machine perspective, algorithms may be perceived to possess some degree of moral agency and responsibility, yet this is likely not the case, as they lack sentience (see Véliz, 2021a). As AI systems are not moral agents, it is unclear who takes responsibility for them (Véliz, 2021b). Hence, if they cause any harm, it is also unclear who is to be held accountable.

From the user's viewpoint, one may tend to ascribe some degree of authority to the outcomes suggested by the machines, as these can be perceived to act independently (Giroux et al., 2022). Such perception of autonomy can lead users to neglect counter-indications or alternative evidence that might contradict the machine-proposed outcomes. Therefore, some users might presume they are not responsible for the governance of the machines, failing to recognise that it is their job to decide the right thing to do with the machine outcomes. These situations can be exacerbated in the case of professionals, or leaders, who pursue their self-interest above all (Wagner, 2021) and in sectors that do not require transparency in decision-making or clear protocols (Ivanov, 2023). Hence, ensuring accountability in AI decision-making represents a serious challenge.

The dark sides in the professional services

The three main issues identified in the previous section are inherent to the use of AI in the professions. The dynamics created by the combination of these issues might lead to disruptions, such as reducing access to services, increasing marginalisation and discrimination, allowing data security breaches and loss of privacy, and the appearance of malicious behaviours, as shown in Figure 1. Having explored these particular dark sides, we draw from the thinking of (Merton, 1968), differentiating between the consequences of the dark side to the primary actor of focus (in our case, the professional and the professions), and the consequences to others (their clients, and the wider society).

Marginalisation, discrimination, and reduced access to services

Many argue that the implementation of AI in professional services fulfils the promise of democratising and increasing opportunities for previously marginalised groups to access goods that before were out of their reach (Susskind & Susskind, 2015). However, despite AI's rapid growth and expansion in services, access can be bounded by usability, pricing, and accountability.

While AI is being adopted in professional firms and professionals are being trained to use AI tools, most of the public is left with a very limited understanding of the new technologies (Seger et al., 2023). This affects the way ordinary people can understand how AI is being used in the provision of services, especially when matters require careful consideration or sophisticated professional advice. This can be worse for specific groups in society that lack access to the internet or have poor digital and general literacy, inhibiting their use or understanding of digital technologies. The more positive views regarding the use of AI tend to ignore the inequalities in access to online resources, and underestimate the

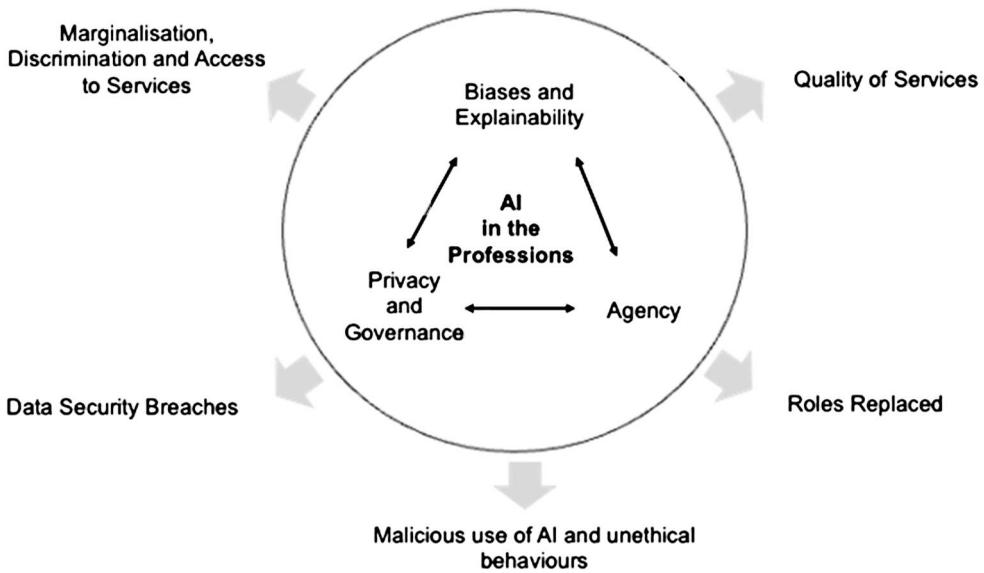


Figure 1. Issues, dynamics, and potential dark sides of AI in the professions.

amount of domain-specific knowledge that technologies and services demand in order to be understood and used. In addition, as explained in the previous section, the ‘black box’ of AI systems can affect transparency in the provision of services or the use of information/data (Véliz, 2021b), and so some consumers, like those described above, can be marginalised.

The use of AI can also affect accountability in the relationship between professionals and clients. This may have noticeable effects on access to public goods such as justice (e.g. when the reasons behind a decision are not sufficiently spelt out, giving rise to uncertainty as to who is responsible for such a decision). AI tools do not always offer full explanations for their recommendations or decisions, blunting the ability for individuals to argue against them or to understand who is accountable for certain outcomes (Fink, 2021).

The evolution of AI, and its nature as a generative technology, might magnify pre-existing biases, producing new classifications and criteria which will increase the potential for new types of biases to emerge. AI is largely powered by data and machine learning algorithms. The data is usually generated or collected through systems created by humans. This means that existing biases in humans can be transferred into the AI systems, or worse, amplified due to complex socio-technical systems that underpin the AI tools. According to Ntoutsis et al. (2020), such transplanted biases might result in discriminatory decisions or outcomes based on features such as race (Angwin et al., 2016), gender (Datta et al., 2014), socioeconomic background, and so forth.

Most significantly for professionals, some AI tools based on biased algorithms may reproduce, or even increase, existing inequalities or wrongful discrimination (Karimi et al., 2018). For example, when AI performs a gate-keeping function such as determining who receives access to support, benefits, etc., and under what conditions (see Johnson

et al., 2019; for the case of access to credit), more discriminatory outcomes are likely to emerge. Similar cases can be found in the insurance industry where the value of the premium, the underwriting of claims, and access to insurance can be influenced by biased parameters (see Loi & Christen, 2021).

Malicious use of AI and unethical behaviours

Although some AI tools can help companies spot unethical behaviour that previously might have been unnoticed (Haenlein et al., 2022), AI also has the potential to be used with purposes different from what it was intended. Illegal use of AI can harm the credibility that stakeholders have in an organisation and consequently its reputation (Belanche et al., 2024). Wagner (2021) explains how AI's capacity for automation and the augmentation of knowledge can exacerbate the dark side of management. In practice, employees and managers in control of AI might be tempted to implement unhealthy managerial practices, such as excessive data surveillance of subordinates' behaviours, micromanagement, and misuse of their authority as controllers of the technology (Wagner, 2021).

Wider access to AI tools brings unscrupulous users to re-purpose technology for abusing others in need of, or looking for, professional advice at a more accessible price (Seger et al., 2023). Increasing progress in AI research and capabilities can lead to potential misuse and misalignment with professional values. Specifically, AI that is developed as open-source (i.e. technology developed by collaboration among users) can present significant threats, because it is more vulnerable to malicious use and modification as it lacks supervision (Ng et al., 2020; Seger et al., 2023). This can result in people with minimal technological experience, or no familiarity with the foundations of technology, using these tools in ignorance of the threat it might pose to their own, and other people's, interests.

AI users may be tempted to exploit these tools to perform actions that violate moral standards and norms, intending to harm an organisation, its members, or other parties (Belanche et al., 2024; Treviño et al., 2006; Wagner, 2021). Cao et al. (2023) suggest that AI can affect how people feel about themselves, that is, how AI becomes part of them, of their knowledge and abilities. This updated identity can induce unethical behaviours insofar as it creates psychological entitlement (e.g. people may sometimes feel rare, unique, valuable, and entitled to treat others with disrespect; Cao et al., 2023; Wagner, 2021).

This can also be seen from a different perspective. Empirical evidence has shown that people are more likely to engage in unethical behaviours when interacting with, or through, AI technology rather than with other humans (Giroux et al., 2022; Kim et al., 2023), as people lack a sense of guilt when interacting with machines. This can lead to immoral behaviours. The underlying idea is that, when using technology, humans may be tempted to behave unethically because it is the machines who are doing the 'dirty' work, thereby partly obviating their agency as users of technology. Some questionable practices can also permeate at the group level. As Bankins et al. (2023) found in their revision of the existing literature, algorithms can be used to control employees (i.e. surveillance), especially those working flexibly and remotely.

Questionable practices and behaviours by employees and leaders can have significant consequences for firms, including effects on their reputation, performance, and their clients' rights. Cases of malicious or unethical use of AI in professional services may

range from surveillance and micromanagement within Professional Service Firms (PSFs) (e.g. Moore, 2019), to email scams and full-scale cyber-warfare attacks (Florida et al., 2018). The effects of these threats can induce significant occupational health and safety risks in the workplace (Moore, 2019).

Quality of services

Artificial Intelligence is being used to interact directly with customers in front-line services, changing how firms establish relationships with consumers (Flavián & Casalo, 2021). Although presented as an innovation, the use of technology can affect customer choices and experience, and, to some degree, be disruptive to what consumers need or expect from service providers (Belanche et al., 2024; Bock et al., 2020). As for other technologies (see Flavián et al., 2024 for the case of the metaverse), users can feel lost when using AI, and without proper guidance, their experience could be hindered, triggering a negative perception of the service or the technology. Therefore, the risks that AI presents for the quality of services can be seen from two angles: disruption in the service as a product of vulnerabilities due to malfunction or lack of supervision, and failure to satisfy customer expectations through the combination of human-machine providers of service.

Florida et al. (2018) argue that the use of AI in skill-intensive services (e.g. healthcare diagnosis or legal provision) can create significant vulnerabilities in the case of malfunction. This is because the use of technologies is subject to risks in different areas, such as development, implementation, and monitoring. Users are not always trained to deal with these vulnerabilities. Indeed, some researchers argue that most companies have limited incentives to rigorously examine the implications of implementing technologies (Katyal, 2019; Morley et al., 2021). In the case of AI, Castillo et al. (2021) found five main antecedents to when AI-powered interactions can fail and lead to loss of value when used: authenticity issues, cognition challenges, affective issues, functionality issues, and integration of conflicts. Most of these problems were found to emerge in the service providers. Yet, interestingly, companies tend to overlook or ignore these antecedents, risking a failure in the provision of services.

Naumov (2019) and Belanche et al. (2024) critically argue that, despite the advantages robots, artificial intelligence, and service automation offer to services, it remains challenging to find the balance between machine and human interaction with consumers. In practice, providing a more standardised service through AI tools does not always align with consumers' expectations (Belanche et al., 2024). When services put the customer and the satisfaction of their desires, expectations, and needs as a priority, the complexity of the interaction between provider and customer increases. For instance, more complex requirements (e.g. legal cases that border the application of the law) will require a more specialised experience and attention to detail that AI tools might not be able to provide. This questions the limits to where and when AI can be used to serve consumers, which, most of the time, are bounded by the training and development of the technology. Whilst AI can doubtless satisfy many consumer expectations and requirements, it is unlikely to meet the needs of all consumers and may even provide some of them with an actively unpleasant experience. Some researchers (Belanche et al., 2024) even argue that the individual feelings about AI (e.g. feeling threatened because it might take over their occupation) can worsen the attitude towards AI-delivered services.

Another significant issue that professional firms experience is related to the accuracy of the services provided. As we know, AI and human intelligence can offer new solutions to old and new problems. Because of this, their potential to be disruptive also entails some risks (Pemer & Werr, 2023) partly shaped by the reduced extent to which humans are in control of the decision-making process (Raisch & Krakowski, 2021). Reducing monitoring in AI technologies decreases the potential for preventing, or addressing, errors that arise or harm users (Floridi et al., 2018). When adopting technology, some companies might focus only on their self-interest when applying ethical principles and establishing monitoring actions (see McMillan & Brown, 2019). In this sense, their concern for ethical aspects, monitoring, and control will be limited by their motivations to obtain benefits or achieve their goals. So far, it has been left to the ethical and moral compass of the individuals, rather than the firms, to recognise that one's self-interest may not align with the goals of society and the public. This calls for firms' action in order to avoid potential damages or obtain more benefits from the service they offer.

Data security breaches

One of the most significant concerns about the use of technology, also prevalent in the use of AI, is the loss of personal privacy. Protecting the consumers' personal data should be a priority in the service industry. However, technological development, especially in AI, has increased concerns about breaches in data privacy. Data breaches are viewed as a violation of trust and a service failure by companies (Chen & Jai, 2021). It happens when an unauthorised entity accesses a firm's caches of personal information (usually through cyber-attacks on a system or network). Organisations do not immediately discover many data breaches (Chen & Jai, 2021), and it can take time until they detect and measure the damage caused by data breaches.

Furthermore, consumers do not realise how much personal data companies possess, and may not understand the implications of a data breach or who to blame (Bansal & Zahedi, 2015). According to Véliz (2021b), because personal data can be very profitable, data privacy has become a significant concern in this technological era. Through the interaction with digital technologies, a vast amount of data is collected and processed by companies, not always for the benefit of the consumers, but for turning data into a commodity or the recursive self-improvement of the technology. When this is done, privacy and data governance come into question, because customers' data expose aspects that they might not be aware companies are using to develop technology or for innovation. People should have complete control over their data, and companies should inform them of their use in any respect (Véliz, 2021b).

Failure to protect people's privacy will affect a company's credibility and service reputation. Using personal data for unauthorised purposes goes against the trust that customers put in their relationships with companies, such as when they consent to the use of their data in loyalty and reward programs. Indeed, Chen and Jai (2021) found that a decrease in trust will be more pronounced where there was a stronger previous relationship with the company rather than where there was not (e.g. loyalty vs non-loyalty program consumers). Failing to protect customers' personal information will, in fact, also affect a company's financial performance and decrease its market value (Martin et al., 2017).

Roles replaced

Although AI is triggering innovation by reshaping services in diverse tasks, some researchers argue that it is also threatening human jobs (Acemoglu, 2021; Floridi et al., 2018; Spring et al., 2022; Susskind & Susskind, 2015). Huang and Rust (2018) developed a theory to address the double-edged impact of AI tools on services. According to them, AI could mainly replace human jobs at the task level rather than the job level. Homogeneous and repetitive tasks are the first to be replaced because AI can easily automate them. For instance, technologies like chatbots have replaced the first point of contact with clients and become the most immediate and ready-to-use advice for services.

The promise of increasing efficiency gained through AI is not exempted from the paradox explained in the 'ironies of automation' (see Bainbridge, 1983). AI systems are still engineered by humans, hence both technical and human factors are still relevant. In this respect, Huang and Rust (2018) argue that tasks that require intuition and empathy will be more difficult to replace. Therefore, although AI can be seen as a threat that leads to higher unemployment and the reconfiguration of roles (Kaplan & Haenlein, 2020), especially for lower-skilled and less-educated workers (Fossen & Sorgner, 2022), the main challenge is presented in the integration of human and machine, and how to preserve efficiency, mitigate errors, and adapt skills. However, new developments in AI may lead to the replacement of humans in *some* high knowledge-intensive tasks in PSFs, which might threaten the work-related identity of many professionals (Cordasco et al., 2021). Belanche et al. (2020) argue that the integration of AI and robotics has the highest potential to replace jobs in service companies in the future, yet the complexity of its implementation and potential ethical concerns require looking at this with caution. The deployment of AI in PSFs will still create tensions, which will undoubtedly affect how employees perceive the technology and their updated roles.

The fiduciary duty: professional identity and trust

The dark sides highlighted above, while unprecedented in magnitude, have always threatened the professions. The key element that underpins such threats is the information asymmetry between professionals and clients, usually shaped by the high knowledge intensity inherent in the professional sectors (Torres, 1991; Von Nordenflycht, 2010). The body of knowledge specific to a particular profession, and over which professionals exercise a certain monopoly, has both formal and informal implications. Formal, to the extent that professionals can, arbitrarily and legally, define who can legitimately use that body of knowledge, mute competition, and regulate the modalities in which such knowledge is used. Informal, in the sense that such a knowledge base is complex and hard to acquire without the formal training required to become a professional, thus informally limiting its use to *insiders*. The formal and informal aspects that give character to the knowledge-intensive professions have a clear and immediate implication in practice, which is the opacity of the quality of the services delivered (Torres, 1991). Therefore, given the formal and informal obstacles to accessing the professional knowledge base, there is a large asymmetry between professionals and clients, which gives rise to a principal-agent (the client and the professional, respectively) problem, which so far, has mainly been addressed through ethical codes of conduct that, in turn, shape professional values and identities.

The staple of ethical codes of conduct to solve the principal-agent problem is rooted in the concept of fiduciary duty (Sealy, 1962). The underlying idea is that, given the knowledge asymmetry between the professional and the client, the professional has a bundle of duties to fulfil, which includes a duty of loyalty in pursuing the objectives of the client and a duty of care in performance (Easterbrook & Fischel, 1993). Such duties are typically enforced through the legal, and extra-contractual, responsibility of the professional to act in the best interest of the client.

However, it would be wrong to imply that it is the legal provisions that solve the principal-agent problem between clients and professionals (Brien, 1998). A much more plausible view is that the legal liability of the professional is just one tool that, along with codes of conduct and standardised educational paths (DiMaggio & Powell, 1983), contributes to the emergence of a professional culture that, consequently, shapes professional values and norms which guide behaviour and elicits trust on the part of the clients and the general public.

In order to clarify this aspect, we may consider the characterisation of professionals as largely motivated by incentives that lie outside the strict economic realm and unfettered self-interest. In doing so, we may look at Scott's work on the cultural-cognitive and normative pillars as key determinants of professional behaviour (Scott, 2008). Alongside this, Abbott (1983) and Van Maanen and Barley (1984) suggest that professionals are largely

influenced by intangible elements such as belonging to a category of highly skilled workers, performing high knowledge-intensive tasks, obtaining recognition for pursuing higher or noble scopes, contributing to societal well-being, and acting in accordance with values and ideals that attach a certain social status to the profession. (Cordasco et al., 2021, p. 5)

These ideas lead to the argument that it is this professional culture, permeated by values and norms, that shapes professionals' intrinsic motivation to pursue higher and noble scopes and that this, in turn, generates trust and solves the principal-agent problem between clients and professionals. It is also the professional culture that prevents, or limits, the emergence of institutional arrangements that could lead to the commoditisation of the professions. For instance, the involvement of non-professionals in the ownership of PSFs is discouraged, when not entirely prohibited, in an effort to defend professional autonomy and avoid the emergence of incentives that may undermine trust between clients and professionals (Von Nordenflycht, 2008).

The same professional culture, along with standardised educational paths, provides guarantees with respect to the quality of the services being offered and sets extra layers of controls, which should limit the deployment and adoption of innovations that, while potentially profit-maximising, might threaten the high-calibre of professional advice and its tailor-made nature. That is, the professional culture, with its values and norms, acts as a filtering mechanism that prevents the *commercial* evolution of professional work and firms, keeping incentives aligned between clients and professionals (Broschak, 2015) and stimulating broader societal trust in professional services, which is a crucial public good (Bagdoniene & Jakstaite, 2009). As seen in Figure 2, the fiduciary duty emerges in the relationship between professionals and beneficiaries, who are embedded in wider structures that produce tensions and determine the boundaries of the professional practice.

A market for lemons in professional services

This section provides a narrative that helps us break down how the promise of AI might unfold in a rather undesirable scenario, at least in the short and medium term. The rationale of such a contribution is that it is only by understanding how the process of AI adoption and diffusion might pan out that we can provide sound policy and strategic advice at various institutional levels (firms, professional bodies, regulators, etc.) in order to prevent the emergence of undesirable states.

We posit that the adoption of AI in professional services, and, especially, the emergence of AI tools creates a selection problem that threatens the overall quality of professional services on offer, thus exposing the public to the various dark sides highlighted above. In particular, two combined key drivers can undermine professional-client relationships, by altering the expectations that are involved in sustaining fiduciary duty: first, a sloppy and uncoordinated process of AI adoption on the PSFs' side; secondly, the widespread emergence of readily available AI tools that can provide pseudo-professional advice to customers (e.g. large language models). Our hypothesis is that the public's awareness of AI deployment in the craft and delivery of professional services, alongside the public's increased familiarity with tools that seemingly provide professional advice, reduces buyers' reserve prices for professional services.

In order to illustrate why this may happen, let us offer a hypothetical exemplar:

Betty typically purchases legal advice and services from Jane's firm. Although Betty lacks technical knowledge of the law, she is happy with the quality of services Jane has provided thus far, which seems always tailored to her needs and circumstances. Jane is willing to experiment with AI in her firm, in an effort to scale up some less knowledge-intensive tasks and augment her professional expertise. However, some of the AI-powered services she rolls out have glitches that Betty is able to notice, and some professional advice becomes more generic and less tailored to Betty's specific circumstances. At the same time, Betty tries to take advantage of AI-powered tools, freely available on the internet, to obtain professional

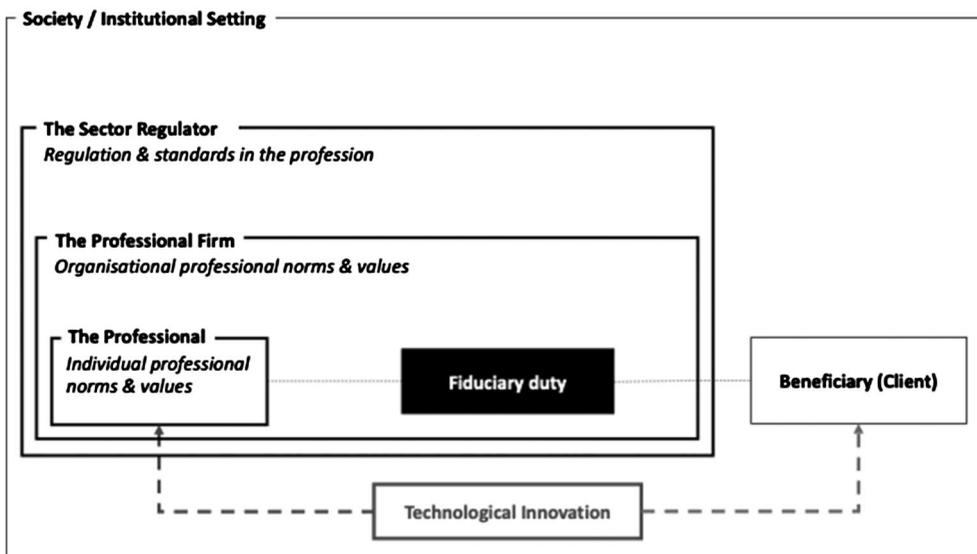


Figure 2. A multilevel perspective situating the fiduciary duty in the professions.

advice. Betty now fails to see the difference between Jane's AI-powered professional advice and the sort of advice she gets on the internet for free. As a result, her reserve price for Jane's services decreases. Jane, in turn, is unwilling to offer tailored professional advice for Betty's reserve price and is thus forced to reduce the quality of her services.

This story resembles nothing but an instance of Akerlof's Market for Lemons (Akerlof, 1970). Information asymmetries create a selection problem that decreases the overall quality of the goods and services being exchanged in a market. In the professions, the problem of asymmetric information is normally solved through trust between clients and professionals, fostered by the fiduciary duty, which we have discussed above. The underlying principle is that the large and specialised body of knowledge warranted by professional work makes it essential to establish trust between clients and professionals, through formal and informal institutional devices.

What is interesting in the example above is that Jane's process of adoption and deployment of AI in her daily practices does not come without challenges that Betty is able to notice. The technological maturity of AI tools does not allow for a smooth transition that makes the adoption and deployment of the new technology invisible to clients. These *glitches* may undermine Betty's trust in Jane, as Betty fails to see Jane as acting with care and in her best interests. At the same time, Betty fails to grasp the difference between Jane's AI-powered services and the sort of professional advice she obtains freely on the internet.

These two aspects combined, through eroding or exterminating trust between clients and professionals, generate the adverse selection problem. The observed glitches can make a client doubt a professional's care in representing their interests and in providing tailored services (Castillo et al., 2021), as well as reduce the client's reserve price for professional services.

One may intuitively point out that, in the example above, Jane has an incentive to adopt and deploy AI only to the extent that Betty perceives it as ultimately safe and tailored to her needs or does not perceive it at all. In this respect, AI adoption should not threaten the fiduciary relationship between clients and professionals. However, Jane's incentives also go in other directions: first, AI is a means to generate economies of scale by automating less knowledge-intensive tasks (Huang & Rust, 2018); second, the race to AI-powered services may reward those who come first through reputational effects; third, some firms may take AI-powered services as an experiment in living (Cordasco, 2023) and make their primary aim that of helping the promise of AI to unfold in services.

These contrasting incentives are likely to give rise to an uncoordinated process of adoption within a profession that may lead to a loss in trust between clients and professionals, not only leading to an overall decreased quality in the offering of professional services, but also sharpening inequalities. The distrust surrounding AI-powered services may lead to a polarisation in the differentiation between a tailored professional services offering, which may become increasingly expensive, and AI-powered professional services readily available to most. These inequalities would betray the promise of AI. Far from widening access to crucial public goods, they would reinforce a state of affairs in which quality services, and the resulting societal goods, are a prerogative of the well-off members of our society.

Reflecting and responding to the dark side

The crucial question then becomes: How do we facilitate a smooth transition to reliable AI-powered professional services? A transition that does not undermine the trust that clients currently place in professionals' compliance with the fiduciary duty, and that does not lead to a state in which services' overall quality is ultimately decreased. Furthermore, the differentiation between human-tailored and AI-powered professional services exacerbates existing inequalities. We must ensure that checks and balances, aimed at avoiding an uncoordinated adoption process, will not create bottlenecks in innovation. Therefore, we propose two main aspects that must be considered: regulation and professional culture.

First, considering both the AI and the professional side requires looking at regulation. General regulation of AI products is one way to work towards ensuring a smooth transition to AI-powered services. This is more than ever salient due to the development of LLMs, whose reliability for non-malicious use is currently being questioned and whose potential for malicious employment has only recently begun to be recognised (Acemoglu, 2021; Véliz, 2021a). In this regard, a smooth transition to AI-powered services would benefit from ensuring products' compliance with the principles and rules identified by many AI ethicists (see Floridi et al., 2018).

The main issue with addressing the transition to AI-powered services through the general regulation of AI products, is that it is such a broad a regulatory space that it will take a wide variety of inputs and, potentially, an unlimited time to tackle. Furthermore, the expansive range of AI products and applications may not lend itself to a unified framework of regulation. Different products may call for different desiderata and requirements, or might demand different degrees of adherence and compliance to the same rules and principles. For instance, it is unclear whether LLMs ought to have the same degree of reliability or transparency in all tasks they perform. Intuitively, trivial requests should not require the same level of due diligence that more sensitive prompts should take.

A much more promising approach would consist of addressing regulation from the standpoint of the professions. The underlying aim would be to define a unified framework and strategy within professional sectors that would design and devise rules facilitating a coordinated and secure process of adoption and deployment of AI-powered services and products. This is by no means easy, as a sector-specific regulatory approach would have to identify an optimal balance between the fostering of competition, by leaving space for innovation, and constraining the deployment of AI in order to ensure quality and safety, thus aiming to preserve the public's trust in the professions.

Furthermore, there is the challenge of ensuring compliance with a sector's regulatory framework, which constrain the roll-out of AI-powered services, as it is not clear how regulatory bodies would manage to enforce such rules, given that the use of AI in professional services can be opaque (Seger et al., 2023). Finally, also linked to compliance, there is the challenge represented by market actors that are looking to use AI to provide professional advice from outside the professional realm, that is, without being licensed. While unlicensed professional advice is formally sanctioned by the law and lacks guarantees, such as legal liability, which sustains the fiduciary duty, the line between professional and pseudo-professional can often be blurry to customers.

After interviewing 49 users of AI across 24 organisations, Kelley (2022) found that a common belief is that AI principles or regulations are not enough to prevent unethical outcomes. The second aspect of our hypothesis on preserving trust in the profession focuses on culture and informal institutions. The underlying idea is that informational asymmetries between professionals and clients had created opportunities and incentives for breaches of fiduciary duty long before the development of AI tools. Yet, professional values and norms have managed to keep professional behaviour in check, promptly isolating cases of misconduct. Whilst there are reasons to believe that AI deployment could represent a major challenge to trust in the professions, it is only countered by fostering an environment that is focussed on the values and norms embedded in Van Maanen and Barley's account of professional culture and in Scott's description of the normative pillar that has characterised the professions thus far.

Our idea is that it is only by cultivating a strong identification amongst professionals with the scope and aims of the professions, and by creating fertile ground for the emergence of value-laden incentives, that we can efficaciously respond to the collective action problem faced in respect of AI adoption and deployment. The question, though, still remains: how do we foster such a culture, and how do we do so without creating unnecessary informal obstacles to AI-driven innovation?

This pushes attention to the education and standards in the professions. One of the prerogatives of the professions, and of the classic professions in particular, is to have a designated and mandatory educational path in order to obtain formal licensing to exercise professional functions (Von Nordenflycht, 2010). Professional and regulatory bodies maintain a certain control over the cultural-cognitive elements of a certain profession, but also over the normative pillar. Such control is a double-edged sword: on the one hand, it helps to preserve professional values by transferring them from the old to the new generation of professionals, thus creating social norms that enforce informal sanctions for misconduct; on the other hand, mandatory and uniform educational paths often fail adequately to accommodate innovation. In this respect, the uncoordinated process of AI adoption observed in certain sectors can be attributed to the lack of flexibility in professional educational paths to accommodate new ways of working, which is leading to a polarisation between 'classical' professionals and tech-savvy ones (e.g. the informal distinction between legal and legal tech firms in law, Trincado-Munoz et al., 2023; van Meeteren et al., 2022). We suggest that efforts should be made to introduce or foster programmes to cultivate AI-related skills in would-be professionals, emphasising ethical considerations and principles within sector-specific educational paths. Not only would this help future professionals in their careers, but it might also smooth the transition towards AI-powered services by helping coordinate the process of adoption and deployment.

Strictly connected to this first aspect is the idea of using *standards* as opposed to formal regulatory devices to help guide a coordinated transition to AI-powered services. Setting standards is a very effective method of channelling the ways in which AI can be adopted and deployed, as standards provide salient solutions to coordination problems that can reduce uncertainty across a sector. We may think of standards as helping to form conventions (Lewis, 2008): when agents A and B are broadly indifferent between option *x* or *y*, but exhibit a preference for coordination (i.e. for both to do *x* or both to do *y*), standards provide them with a salient solution to the coordination problem. For

instance, standards may help firms coordinate safety, accuracy, and data management thresholds.

Standards do not help solely by providing salient solutions to coordination games, but they also serve as a way to create social norms that transform uncooperative games into coordination games (Bicchieri, 1990). Sometimes, indeed, standards identify coordination points that some may see as ultimately disadvantageous, despite being conducive to an overall collective good. In such cases, ensuring compliance becomes problematic as agents have an incentive to defect. However, standards may also help give rise to the emergence of informal sanctions for those who fail to comply, thus de facto enforcing compliance by making it increasingly costly for agents to defect. For instance, a firm may benefit from standards that define a more-or-less lax threshold for data security or accuracy in AI-powered services, to the point where, if the coordination point indicated by the standards is far from one's wishes, there is a high incentive to avoid complying. Yet, informal sanctions associated with defections might make non-compliance costly, thus transforming an uncooperative game into a coordination game.

Education and standards work in tandem to help smooth the transition to AI-powered services, as it is through education that standards can be embedded into the cultural-cognitive and normative pillars, thus becoming part of professional identity. In this regard, our idea is that any attempt to facilitate a coordinated transition to AI-powered services cannot look at regulation, education, and standards in discrete compartments but rather requires a holistic approach that takes into account how these three elements interact.

Concluding, we would like to stress that this paper forges the way for further analysis of the dark sides of AI both in Professional Services, and more widely. In particular, we wish to highlight four main aspects that call for further research. First, although we outline the three foundational issues that can lead to dark sides in the professions, such dark sides are heavily context-dependent, and so is their salience; in this respect, we encourage further empirical research aimed at identifying sector-specific challenges that AI and its uncoordinated deployment could bring about. Second, sector-specific research is needed in order to identify standards that represent an optimal balance between the freedom to innovate and the provision of sound thresholds of safety and accuracy. Third, strictly linked to the above point, sector-specific research should also focus on the customers' experience in interacting with AI. Particularly in an effort to understand how customers' negative experiences, along with the effects of readily available AI offerings of pseudo-professional advice, may affect customers' behaviour. Fourth, sector-specific research should look at how to re-imagine professional education and training that would, on the one hand, cultivate the values underpinning professional identity and, on the other, help accelerate the adoption and use of technical innovations.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by The UK's Economic and Social Research Council (ESRC): Innovating Next Generation Services Through Collaborative Design [Project ES/S010475/1].

References

- Abbott, A. (1983). Professional ethics. *American Journal of Sociology*, 88(5), 855–885. <https://doi.org/10.1086/227762>
- Acemoglu, D. (2021). *Harms of AI* (No. w29247). National Bureau of Economic Research.
- Acemoglu, D., & Restrepo, P. (2018). Artificial intelligence, automation, and work. In *The economics of artificial intelligence: An agenda* (pp. 197–236). NBER.
- Akerlof, G. A. (1970). The market for “Lemons”: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3), 488–500. <https://doi.org/10.2307/1879431>
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias. Retrieved September 14, 2023, from <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Bagdoniene, L., & Jakstaite, R. (2009). Trust as basis for development of relationships between professional service providers and their clients. *Economics and Management*, 14(14), 360–366.
- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775–779.
- Bankins, S., Ocampo, A. C., Marrone, M., Restubog, S. L. D., & Woo, S. E. (2023). A multilevel review of artificial intelligence in organizations: Implications for organizational behavior research and practice. *Journal of Organizational Behavior*, 45(2), 159–182.
- Bansal, G., & Zahedi, F. M. (2015). Trust violation and repair: The information privacy perspective. *Decision Support Systems*, 71, 62–77. <https://doi.org/10.1016/j.dss.2015.01.009>
- Belanche, D., Belk, R. W., Casalo, L. V., & Flavián, C. (2024). The dark side of artificial intelligence in services. *The Service Industries Journal*, 44(3–4), 149–172. <https://doi.org/10.1080/02642069.2024.2305451>
- Belanche, D., Casalo, L., Flavián, C., & Schepers, J. (2020). Service robot implementation: A theoretical framework and research agenda. *The Service Industries Journal*, 40(3–4), 203–225. <https://doi.org/10.1080/02642069.2019.1672666>
- Belk, R. (2021). Ethical issues in service robotics and artificial intelligence. *The Service Industries Journal*, 41(13–14), 860–876. <https://doi.org/10.1080/02642069.2020.1727892>
- Belk, R. W., Belanche, D., & Flavián, C. (2023). Key concepts in artificial intelligence and technologies 4.0 in services. *Service Business*, 17, 1–9. <https://doi.org/10.1007/s11628-023-00528-w>
- Bicchieri, C. (1990). Norms of cooperation. *Ethics*, 100(4), 838–861. <https://doi.org/10.1086/293237>
- Bock, D. E., Wolter, J. S., & Ferrell, O. C. (2020). Artificial intelligence: Disrupting what we know about services. *Journal of Services Marketing*, 34(3), 317–334. <https://doi.org/10.1108/JSM-01-2019-0047>
- Brien, A. (1998). Professional ethics and the culture of trust. *Journal of Business Ethics*, 17, 391–409. <https://doi.org/10.1023/A:1005766631092>
- Brooks, C., Gherhes, C., & Vorley, T. (2020). Artificial intelligence in the legal sector: Pressures and challenges of transformation. *Cambridge Journal of Regions, Economy and Society*, 13(1), 135–152. <https://doi.org/10.1093/cjres/rsz026>
- Broschak, J. P. (2015). Client relationships in professional service firms. In L. Empson, D. Muzio, J. Broschak, & B. Hinings (Eds.), *The Oxford handbook of professional service firm* (pp. 304–326).
- Cao, L., Chen, C., Dong, X., Wang, M., & Qin, X. (2023). The dark side of AI identity: Investigating when and why AI identity entitles unethical behavior. *Computers in Human Behavior*, 143, Article 107669. <https://doi.org/10.1016/j.chb.2023.107669>
- Castillo, D., Canhoto, A. I., & Said, E. (2021). The dark side of AI-powered service interactions: Exploring the process of co-destruction from the customer perspective. *The Service Industries Journal*, 41(13–14), 900–925. <https://doi.org/10.1080/02642069.2020.1787993>
- Chen, H. S., & Jai, T. M. (2021). Trust fall: Data breach perceptions from loyalty and non-loyalty customers. *The Service Industries Journal*, 41(13–14), 947–963. <https://doi.org/10.1080/02642069.2019.1603296>
- Cordasco, C., Gherhes, C., Brooks, C., & Vorley, T. (2021). An institutional 23 taxonomy of adoption of innovation in the classic professions. *Technovation*, 107, Article 102272. <https://doi.org/10.1016/j.technovation.2021.102272>
- Cordasco, C. L. (2023). The ethics of entrepreneurship: A millian approach. *Journal of Business Ethics*, 1–13. In press.

- Datta, A., Tschantz, M. C., & Datta, A. (2014). *Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination*. arXiv preprint arXiv:1408.6491.
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147–160. <https://doi.org/10.2307/2095101>
- Easterbrook, F. H., & Fischel, D. R. (1993). Contract and fiduciary duty. *The Journal of Law and Economics*, 36(1, Part 2), 425–446. <https://doi.org/10.1086/467282>
- Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, 14(1), 57–74. <https://doi.org/10.2307/258191>
- Faulconbridge, J., Sarwar, A., & Spring, M. (2023). How professionals adapt to artificial intelligence: The role of intertwined boundary work. *Journal of Management Studies*. In press.
- Fink, M. (2021). The EU artificial intelligence act and access to justice. *EU Law Live*. Retrieved from <https://scholarlypublications.universiteitleiden.nl/access/item>
- Flavián, C., & Casaló, L. V. (2021). Artificial intelligence in services: Current trends, benefits and challenges. *The Service Industries Journal*, 41(13–14), 853–859. <https://doi.org/10.1080/02642069.2021.1989177>
- Flavián, C., Ibáñez-Sánchez, S., Orús, C., & Barta, S. (2024). The dark side of the metaverse: The role of gamification in event virtualization. *International Journal of Information Management*, 75, 102726. <https://doi.org/10.1016/j.ijinfomgt.2023.102726>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., & Schafer, B. (2018). AI4People: Twenty recommendations for an ethical framework for a good AI society. *Minds and Machines*, 28, 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Fossen, F. M., & Sorgner, A. (2022). New digital technologies and heterogeneous wage and employment dynamics in the United States: Evidence from individual-level data. *Technological Forecasting and Social Change*, 175, Article 121381. <https://doi.org/10.1016/j.techfore.2021.121381>
- Giroux, M., Kim, J., Lee, J. C., & Park, J. (2022). Artificial intelligence and declined guilt: Retailing morality comparison between human and AI. *Journal of Business Ethics*, 178(4), 1027–1041. <https://doi.org/10.1007/s10551-022-05056-7>
- Haenlein, M., Huang, M. H., & Kaplan, A. (2022). Guest editorial: Business ethics in the era of artificial intelligence. *Journal of Business Ethics*, 178(4), 867–869. <https://doi.org/10.1007/s10551-022-05060-x>
- Huang, M. H., & Rust, R. T. (2018). Artificial intelligence in service. *Journal of Service Research*, 21(2), 155–172. <https://doi.org/10.1177/1094670517752459>
- Ivanov, S. (2023). The dark side of artificial intelligence in higher education. *The Service Industries Journal*, 43(15–16), 1055–1082. <https://doi.org/10.1080/02642069.2023.2258799>
- Jin, G. Z. (2018). Artificial intelligence and consumer privacy. In A. Agrawal, J. Gans, & A. Goldfarb (Eds.), *The economics of artificial intelligence: An agenda* (pp. 439–462). University of Chicago Press.
- Johnson, K., Pasquale, F., & Chapman, J. (2019). Artificial intelligence, machine learning, and bias in finance: Toward responsible innovation. *Fordham Law Review*, 88, 499.
- Kaplan, A., & Haenlein, M. (2020). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. *Business Horizons*, 63(1), 37–50. <https://doi.org/10.1016/j.bushor.2019.09.003>
- Karimi, F., Génois, M., Wagner, C., Singer, P., & Strohmaier, M. (2018). Homophily influences ranking of minorities in social networks. *Scientific Reports*, 8(1), 11077. <https://doi.org/10.1038/s41598-018-29405-7>
- Katyal, S. K. (2019). Private accountability in the age of artificial intelligence. *UCLA Law Review*, 66, 54.
- Kelley, S. (2022). Employee perceptions of the effective adoption of AI principles. *Journal of Business Ethics*, 178(4), 871–893. <https://doi.org/10.1007/s10551-022-05051-y>
- Kim, T., Lee, H., Kim, M. Y., Kim, S., & Duhachek, A. (2023). AI increases unethical consumer behavior due to reduced anticipatory guilt. *Journal of the Academy of Marketing Science*, 51(4), 785–801. <https://doi.org/10.1007/s11747-021-00832-9>

- Leitner-Hanetseder, S., Lehner, O. M., Eisl, C., & Forstenlechner, C. (2021). A profession in transition: Actors, tasks and roles in AI-based accounting. *Journal of Applied Accounting Research*, 22(3), 539–556. <https://doi.org/10.1108/JAAR-10-2020-0201>
- Lewis, D. (2008). *Convention: A philosophical study*. John Wiley and Sons.
- Loi, M., & Christen, M. (2021). Choosing how to discriminate: Navigating ethical trade-offs in fair algorithmic design for the insurance sector. *Philosophy & Technology*, 34, 967–992. <https://doi.org/10.1007/s13347-021-00444-9>
- Martin, K. D., Borah, A., & Palmatier, R. W. (2017). Data privacy: Effects on customer and firm performance. *Journal of Marketing*, 81(1), 36–58. <https://doi.org/10.1509/jm.15.0497>
- McMillan, D., & Brown, B. (2019, November 19–20). Against ethical AI: Guidelines and self interest. In *Proceedings of the Halfway to the Future Symposium 2019 (HTFF 2019)* (p. 3). ACM. <https://doi.org/10.1145/3363384.3363393>
- Merken, S. (2023). *New York lawyers sanctioned for using fake ChatGPT cases in legal brief*, Reuters. Thomson Reuters. Retrieved September 26, 2023, from <https://www.reuters.com/legal/new-york-lawyers-sanctioned-using-fake-chatgpt-cases-legal-brief-2023-06-22/>.
- Merton, R. K. (1968). *Social theory and social structure*. Simon and Schuster.
- Moore, P. V. (2019). OSH and the future of work: Benefits and risks of artificial intelligence tools in workplaces. In *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Body and Motion: 10th International Conference, DHM 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part I 21* (pp. 292–315). Springer International Publishing.
- Morley, J., Elhalal, A., Garcia, F., Kinsey, L., Mökander, J., & Floridi, L. (2021). Ethics as a service: A pragmatic operationalisation of AI ethics. *Minds and Machines*, 31(2), 239–256. <https://doi.org/10.1007/s11023-021-09563-w>
- Naumov, N. (2019). The impact of robots, artificial intelligence, and service automation on service quality and service experience in hospitality. In *Robots, artificial intelligence, and service automation in travel, tourism and hospitality* (pp. 123–133). Emerald Publishing Limited.
- Ng, K., Niven, K., & Hoel, H. (2020). ‘I could help, but...’: A dynamic sensemaking model of workplace bullying bystanders. *Human Relations*, 73(12), 1718–1746. <https://doi.org/10.1177/0018726719884617>
- North, D. C. (1991). Institutions. *Journal of Economic Perspectives*, 5(1), 97–112. <https://doi.org/10.1257/jep.5.1.97>
- Ntoutsis, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdli, W., Vidal, M. E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., & Kompatsiaris, I. (2020). Bias in data-driven artificial intelligence systems – an introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(3), e1356.
- Park, S. S., Tung, C. D., & Lee, H. (2021). The adoption of AI service robots: A comparison between credence and experience service settings. *Psychology and Marketing*, 38(4), 691–703. <https://doi.org/10.1002/mar.v38.4>
- Pemer, F., & Werr, A. (2023). Defusing digital disruption through creative accumulation: Technology-Induced innovation in professional service firms. *Journal of Management Studies*, 1–46. In press.
- Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review*, 46(1), 192–210. <https://doi.org/10.5465/amr.2018.0072>
- Scott, W. R. (1994). Law and organizations. In S. B. Sitkin & R. J. Bies (Eds.), *The legalistic organization* (pp. 3–18). Sage.
- Scott, W. R. (2008). Lords of the dance: Professionals as institutional agents. *Organization Studies*, 29(2), 219–238. <https://doi.org/10.1177/0170840607088151>
- Sealy, L. S. (1962). Fiduciary relationships. *The Cambridge Law Journal*, 20(1), 69–81. <https://doi.org/10.1017/S0008197300086943>
- Seger, E., Ovadya, A., Garfinkel, B., Siddarth, D., & Dafoe, A. (2023). *Democratising AI: Multiple meanings, goals, and methods*. arXiv preprint arXiv:2303.12642.
- Spring, M., Faulconbridge, J., & Sarwar, A. (2022). How information technology automates and augments processes: Insights from Artificial-Intelligencebased systems in professional service operations. *Journal of Operations Management*, 68(6–7), 592–618. <https://doi.org/10.1002/joom.v68.6-7>

- Susskind, R. E., & Susskind, D. (2015). *The future of the professions: How technology will transform the work of human experts*. Oxford University Press.
- Torres, D. (1991). What, if anything, is professionalism? Institutions and the problem of change. *Research in the Sociology of Organizations and Occupations*, 8, 43–68.
- Trappenberg, T. P. (2019). *Fundamentals of machine learning*. Oxford University Press.
- Tredinnick, L. (2017). Artificial intelligence and professional roles. *Business Information Review*, 34(1), 37–41. <https://doi.org/10.1177/0266382117692621>
- Treviño, L. K., Weaver, G. R., & Reynolds, S. J. (2006). Behavioral ethics in organizations: A review. *Journal of Management*, 32(6), 951–990. <https://doi.org/10.1177/0149206306294258>
- Trincado-Munoz, F., van Meeteren, M., Rubin, T. H., & Vorley, T. (2023). Digital transformation in the world city networks' advanced producer services complex: A technology space analysis. *Geoforum*, Article 103721. In press. <https://doi.org/10.1016/j.geoforum.2023.103721>
- Van Maanen, J., & Barley, S. R. (1984). Occupational communities: Culture and control in organizations. *Research in Organizational Behavior*, 6(1), 287–365.
- van Meeteren, M., Trincado-Munoz, F., Rubin, T. H., & Vorley, T. (2022). Rethinking the digital transformation in knowledge-intensive services: A technology space analysis. *Technological Forecasting and Social Change*, 179, Article 121631. <https://doi.org/10.1016/j.techfore.2022.121631>
- Véliz, C. (2019). Three things digital ethics can learn from medical ethics. *Nature Electronics*, 2(8), 316–318. <https://doi.org/10.1038/s41928-019-0294-2>
- Véliz, C. (2021a). Moral zombies: Why algorithms are not moral agents. *AI & Society*, 36, 487–497. <https://doi.org/10.1007/s00146-021-01189-x>
- Véliz, C. (2021b). *Privacy is power*. Melville House.
- Von Nordenflycht, A. (2008). The demise of the professional partnership? The emergence and diffusion of publicly-traded professional service firms. In *Future of the Global Law Firm' Symposium* (Vol. 17). Georgetown University Center for the Study of the Legal Profession.
- Von Nordenflycht, A. (2010). What is a professional service firm? Toward a theory and taxonomy of knowledge-intensive firms. *Academy of Management Review*, 35(1), 155–174.
- Wagner, D. (2021). Artificial intelligence and the dark side of management. *ROBONOMICS: The Journal of the Automated Economy*, 1, 10–10.