Philosophy & Technology (2025) 38:60
https://doi.org/10.1007/s13347-025-00890-9

RESEARCH ARTICLE

The Potential and Limitations of Artificial Colleagues

Friedemann Bieber[1,2] · **Charlotte Franziska Unruh**[3]

Received: 3 January 2025 / Accepted: 12 April 2025

Abstract

This article assesses the potential of artificial colleagues to help us realise the goods of collegial relationships and discusses its practical implications. In speaking of artificial colleagues, it refers to AI-based agential systems in the workplace. The article proceeds in three steps. First, it develops a comprehensive account of the goods of collegial relationships. It argues that, in addition to goods at the individual level, collegial relationships can provide valuable goods at the social level. Second, it argues that artificial colleagues are limited in their capacity to realise the goods of collegial relationships: at the individual level, they can at best realise some such goods, and at the social level, they can at best support their realisation. This contradicts Nyholm and Smids' (2020) claim that robots can be good colleagues. The article traces these limitations to particular features of artificial colleagues and discusses to what extent they would hold for radically advanced systems. Third, the article examines the policy implications of these findings. It highlights how the introduction of artificial colleagues, in addition to potentially crowding out human colleagues, will likely impact relations among human colleagues. And it proposes a governance principle that gives strict priority to human collegial relationships.

Keywords Artificial Colleagues · Collegiality · Relationships · Robot Ethics · Work · Human-robot-interaction

✉ Charlotte Franziska Unruh
c.unruh@soton.ac.uk

Friedemann Bieber
friedemann.bieber@philos.uzh.ch

[1] University of Zurich, Zurich, Switzerland

[2] University of Groningen, Groningen, The Netherlands

[3] University of Southampton, Southampton, United Kingdom

Published online: 02 May 2025

1 Introduction

Recent years have seen a rapid rise and proliferation of artificial intelligence (AI) technology. We interact with AI via screen interfaces, but also in 'embodied' form in robots and vehicles. Increasingly, AI does not only shape or mediate our interactions with other human beings but has itself become a (somewhat) independent party to interactions. One important area of application is the workplace.

What we call *artificial colleagues* are AI-based agents– systems that can formulate plans and pursue actions– that operate in the workplace and interact with workers.[1] Examples of artificial colleagues are robots that fetch parcels from warehouses, assistant systems that can make calendar appointments, and chatbots that help workers reason through problems. The prospect of a future where we increasingly interact with artificial colleagues raises the question: will the relationships that workers are able to form and maintain with artificial colleagues differ from the relationships they have with human workers? This question matters because artificial colleagues are likely to become ubiquitous in many professions. With the rise of AI agents, there is now increasing agreement, both among industry leaders and academic experts, that AI will become prominent in the workplace. Consider, for instance, the prediction by OpenAI's CEO that 'in 2025, we may see the first AI agents "join the workforce" and materially change the output of companies' (Altman 2025) or Geoffrey Hinton's prediction that the impact of AI systems on labour markets will be so extensive that a universal basic income will be needed to counterbalance the effects on economic inequality (BBC 2024). We take no stance on whether artificial agents are set to replace human workers at large scale,[2] but even if this were to happen, ethical and political questions regarding artificial colleagues would remain relevant– for the transition period, and for any workplaces in which human and artificial workers would continue to interact.

This article seeks to offer a nuanced assessment of the potential and limitations of artificial colleagues and to assess some policy implications. We argue that AI-based systems can at most *realise some* of the goods of collegial relationships at the *individual level*, and can at best *support*, but might actually hinder, the realisation of the goods of collegial relationships at the *social level*. This contradicts Sven Nyholm and Jilles Smids' (2020) claim that embodied AI-systems can be good colleagues.[3]

This article proceeds in three steps. Section 2 sets out a comprehensive account of the goods of collegial relationships. In so doing, it introduces a distinction between two dimensions: a personal dimension, comprising goods that accrue to parties of the relationships, and a social dimension, comprising goods that constitute social externalities. Drawing on existing work, we distinguish three types of collegial rela-

[1] Gabriel et al. define 'advanced AI assistants […] as artificial agents with natural language interfaces, the function of which is to plan and execute sequences of actions on the user's behalf– across one or more domains– and in line with the user's expectations' (2024, p. 3). Artificial colleagues can be AI assistants, but they need not be: we do not require such systems to have natural language interfaces.

[2] For the claim that replacement is likely, see e.g. Susskind (2020) and Amodei (2024). For a more sceptical view, see, e.g., this recent Brooking Institute report by economists Fleming, Li and Thompson (2024).

[3] Nyholm and Smids offer an argument about robots. As we note in section 3, we believe that embodiment can be conducive to realising collegial relationship goods but is neither necessary nor sufficient.

tionships and identify the individual goods they can generate. We then argue that collegial relationships can produce two social goods: empowerment and social cohesion. Section 3 assesses to what extent relationships with artificial colleagues can realise these goods. We argue that most can only be fully realised in relationships with human colleagues. For each good, we provide a detailed account of why the capacity of artificial colleagues to provide it is limited. Our main argument relies on the claim that artificial colleagues lack robust autonomy, an inner life, and equal moral and political status, but we also entertain the possibility of advanced AI systems that are no longer lacking in these regards. Section 4 briefly explores the policy implications of our findings. It first argues that collegial relationships are integral to the context of work, so even if we could have other kinds of valuable workplace relationships with AI systems, their inability to realise the specific goods of *collegial* relationships would remain an important limitation.[4] It then discusses several policies for regulating artificial colleagues and advocates a specific governance principle, the *Principle of Strict Priority of Human Collegial Relationships*. Section 5 concludes.

Our article has two main upshots. First, *pace* a more optimistic assessment by Nyholm and Smids (2020), we cannot (fully) replace human colleagues with artificial colleagues without sacrificing the opportunity to realise valuable collegial relationships goods. Second, in so far as policy is concerned, the primary focus should not be on whether artificial colleagues can be 'good colleagues,' but rather on whether the introduction of artificial colleagues has detrimental effects on collegial relationships among humans.

2 The Goods of Collegial Relationships

Paradigmatically, colleagues are those who work in similar positions, at similar levels of seniority, and for the same employer.[5] But the term 'colleagues' can be used more broadly. For example, philosophers can be colleagues even when they work at different universities. In fact, two people can be colleagues even without knowing of each other. So, being colleagues is necessary, but not sufficient, for having a collegial *relationship*. Following Seglow (2013), we understand relationships as 'enduring, substantive, mutually affirmed interactions between two or more people' (2013, p. 28). So, *collegial relationships* describe interactions between colleagues (Betzler & Löschke, 2021, p. 218).

In this section, we provide a comprehensive account of the *goods* that can be realised in collegial relationships if they go well.[6] In doing so, we draw on recent philosophical contributions on collegial relationships (Betzler & Löschke, 2021;

[4] We are grateful to an anonymous reviewer for pressing us on this point.

[5] Betzler and Löschke offer the following definition: 'two people qualify as colleagues if they share: (i) the same work content or domain of activity; (ii) the same institutional affiliation or common purpose; and/or (iii) the same status or level of responsibility' (2021, p. 217). For critical discussion, see Bieber and Unruh (forthcoming).

[6] As we acknowledge below, when they do not go well, collegial relationships can generate various bads.

Mlonyeni, 2023, Bieber & Unruh, forthcoming). But we propose a new framing, distinguishing between *individual* and *social* relationship goods

2.1 Individual Vs. Social Relationship Goods

Interpersonal relationships, like friendships and romantic relationships, are of vital importance to the lives of most people (Gheaus, 2022), and the goods realised in these relationships are the subject of lively philosophical debates. However, there are two limitations to this debate. First, in contrast to family relationships (e.g., Brighouse & Swift, 2014), friendships (e.g., Thomas, 1987), and love relationships (e.g., Velleman, 1999), *collegial relationships* have attracted little philosophical attention.[7] Yet, many of us spend a significant part of our lives at work, and often in the company of colleagues, so collegial relationships are a significant potential source of value and well-being in our lives and deserve scrutiny. Second, in philosophical discussions of interpersonal relationships, the focus is typically on *individual* or *personal goods*, such as intimacy, feelings of belonging, or social respect, i.e., on goods that these relationships can provide to the individuals who form them (Gheaus, 2022). But, as Helm (2023) argues for friendships, interpersonal relationships can also have *social* value: their existence can have positive externalities, benefitting communities and society-at-large. The experience of friendship, for instance, can increase people's willingness to act beneficently and change their moral outlooks (see also Friedman, 1989, pp. 6–7). Elizabeth Telfer concurs that 'even those who have no friends are […] better off than they would be if there were no such thing as friendship, since the understanding developed by it and the mutual criticism involved in it will improve the way friends deal with people outside the relationship' (1971, p. 238).[8] We believe that this point generalises: relationships with friends, neighbours and colleagues are a crucial element of a good life; but they can also have valuable social externalities.

The following examination of the goods of collegial relationships is thus structured along this distinction between individual and social goods.

2.2 Individual Goods in Collegial Relationships

Monika Betzler and Jörg Löschke (2021) offer what is, to our knowledge, the first targeted philosophical account of the nature and value of collegial relationships. They claim that collegial relationships ideally take the form of what one might call *professional relationships*– relationships that are rooted in the workplace context and remain somewhat detached. According to Betzler and Löschke (2021), colleagues can obtain two distinct goods in such professional relationships. First, solidarity: colleagues show solidarity by supporting each other and by giving work-related advice (2021, pp. 219–21). This, they claim, can be of particular instrumental value (shared experience and knowledge can render the help of colleagues especially effective (2021, p. 220) and non-instrumental value (by showing solidarity, a colleague

[7] Notable exceptions are Betzler and Löschke (2021), Mlonyeni (2023), and Nyholm and Smids (2020).

[8] Romantic relationships and friendships can also have negative social value, for example by giving rise to cronyism and undue partiality (Lintott, 2015; Thomas, 1999).

'makes the ends of another person in work-related matters her own ends' (Betzler & Löschke, 2021, p. 221). Second, recognition: colleagues can provide recognition by complimenting each other on work achievements and by acknowledging the other's skills and efforts (Betzler & Löschke, 2021, pp. 221–23). According to Betzler and Löschke, collegial recognition has special instrumental value, because colleagues, by way of having an intimate knowledge of the challenges and rewarding aspects of the work, can best judge achievements (2021, pp. 222–23). In addition, they claim, collegial recognition is non-instrumentally valuable, because the understanding that colleagues have for each other helps to validate the other's struggles and feelings (Betzler & Löschke, 2021, p. 223).

Mlonyeni (2023) argues that this account of the goods of collegial relationships is incomplete. According to Mlonyeni, one can be a good colleague not only by being a detached professional, but also by being a *collegial friend.* Mlonyeni describes two goods that collegial friendships can provide. First, emotional involvement and support: collegial friends will be there to provide feedback and advice, to listen, to vent, and to commiserate about work-related matters (Mlonyeni, 2023, p. 114, 123). What distinguishes this support from solidarity among professional colleagues is that collegial friends provide support in light of 'an emotional involvement directed at certain features of the other', such as their personal histories and idiosyncrasies (Mlonyeni, 2023, p. 118). Second, appreciation: friends appreciate the other as a specific person and not just as the occupant of a role, like that of neighbour, rower, or housemate (Mlonyeni, 2023, p. 118). Such appreciation, Mlonyeni claims, is not merely of instrumental, but also of non-instrumental value: friendships, including collegial friendships, are an important part of a meaningful and fulfilling life (2023, p. 119).

As we argue at length elsewhere (Bieber & Unruh, forthcoming), we believe that there is an important third type of collegial relationship: *collaborative relationships*. In collaborative relationships, we act together with others to put our joint intentions into action. For example, two software developers might share an intended end (delivering a product) and act to bring it about (specifying requirements, writing code, testing the product). Collaborative relationships among colleagues can yield two distinct goods.[9] First, shared experience: in collaboratively designing a flower arrangement, florists will puzzle over the colour scheme, discuss which seasonal flowers match the venue, and go through the emotional ups and downs involved in the project together. Their experience will differ from that of completing a solo project, and might include a feeling of camaraderie, of 'being-in-this-together', of rooting for shared success. Shared experience can be instrumentally valuable: it can help us to learn from others, boost morale and motivation, and increase feelings of belonging and connectedness. In virtue of providing meaningful social connection, it can also be of non-instrumental value. Second, collaborative relationships can yield the good of shared achievement. Following Bradford (2015), we understand achievements as valuable outputs that agents bring about in a competent, not-merely-lucky way, where an output can be a product (a completed house) or a process itself (building the house). Shared achievements can be particularly valuable both because collaboration enables larger

[9] The remainder of this paragraph summarises an argument developed in Bieber and Unruh (forthcoming).

projects (and thus weightier outcomes) and because, given the coordination required, successfully managing collaborative work can itself constitute an achievement.

In sum, we submit that collegial relationships come in at least three different forms, each of which offers distinct goods: professional collegial relationships can yield the goods of solidarity and recognition; collegial friendships can generate the goods of emotional support and appreciation; and collaborative collegial relationships can yield the goods of shared experience and shared achievements.

2.3 Social Goods in Collegial Relationships

In addition, collegial relationships can provide social value. Recall Telfer's point that friendships benefit even those without friends, since friendship 'promotes the general happiness by providing a degree and kind of consideration for others' welfare which cannot exist outside it' (Telfer, 1971, p. 238) and by supporting the development of skills in interacting with others, which benefits everyone with whom those who have friends interact (Telfer, 1971, p. 238). We submit that an analogous argument holds for good collegial relationships.

Interacting with people *in* the workplace, we believe, can have a formative effect on our interactions with people *outside* of the workplace. If successful collegial relationships yield the goods of solidarity and recognition, emotional support and appreciation, and shared experience and achievement at the individual level, then it is plausible that they support the broader development of interpersonal skills, such as the ability to effectively communicate, to compromise, to be attentive to the needs of other people, and to find ways to reach mutually beneficial agreements. It is highly plausible that the development of such skills will benefit interactions outside the workplace: we do not, after all, become a different person upon clocking out of work. In fact, obtaining the aforementioned skills in the context of the workplace, rather than simply in interactions with friends, will likely have a stronger impact on how we interact within society: because we are less close to, and thus less biased in favour of, (most of) our colleagues, how we interact with them more closely resembles, and thus plausibly shapes, how we interact with members of the general public. So, if our claims about the individual goods of collegial relationships hold, then it is highly plausible that that they can have positive social externalities.

This line of reasoning is not, of course, entirely new. Estlund (2003), for example, has prominently argued that personal relationships at work can help us build a basis of trust and understanding that underpins our living together.

But in what ways, specifically, does society stand to benefit from successful collegial relationships? Building on our discussion above, we propose that the social goods of collegial relationships primarily come in two forms. First, collegial relationships can help to practice skills that are crucial to our ability to act as democratic citizens, providing the good of *empowerment*: they can foster one's ability to pursue one's goals in the political realm. The relevant skills include, but are not limited to, communication abilities, the capacity to compromise and collaborate, and the aptitude to pursue tasks independently or collectively and to recognise the connec-

tion between specific tasks and the broader production process.[10] These competences matter because they are autonomy-enhancing: knowing how to effectively communicate, how to effectively achieve goals, and how one's actions play out in complex multi-agent situations all help us attain stronger agency, empowering us to effectively and independently pursue our goals. Collegial relationships are one important ground for honing such skills. This is especially true of collaborative collegial relationships: many larger goals cannot be attained by individuals by themselves, so the experience of effectively engaging with others is of particular value.

Second, collegial relationships can contribute to the good of *social cohesion*. In bringing members of society together to interact and share experiences, collegial relationships can help to support feelings of belonging and identification among fellow citizens. Estlund describes how citizens from different backgrounds come together at the workplace, claiming that 'being part of the same organization and getting the job done together tend to create common ground and to cultivate mutual affinity' (Estlund, 2003, p. 25). These relationships, Estlund argues, are of crucial importance in democratic societies: 'The sheer amount of sociability and cooperation that takes place every day in workplaces should place them at the center of any account of what holds a complex, modern democratic society together' (Estlund, 2003, p. 34). Interaction with people of different socio-economic backgrounds, political views, cultural identities, and so on, will plausibly increase familiarity with those with whom one might not have normally interacted, in a context where the shared focus is on the joint work task and on getting it done together– an environment that might help those who do not normally interact to become acquaintances, good colleagues, or even friends. Given the large amount of time that most people spend at work, this is an important context for such interaction. While not all workplaces are equally diverse, those that are can arguably promote social cohesion to a greater extent than more homogenous workplaces.

In emphasising collegial relationships, our account complements the existing literature on the social and political implications of the organisation of work. Schwartz (1982) for example argues that repetitive, low-skilled work undermines worker autonomy; and that in doing so, it prevents workers from developing and exercising the skills they need as democratic citizens (see also Roessler (2012)). Caleb Althorpe, meanwhile, argues that because 'the opportunity to engage in meaningful work [is] a social basis of self-respect' (2022, p. 1), the opportunity to manifestly make a social contribution in one's work is critical for a democratic society. If we are correct, then the nature of collegial relationships is an important further consideration here.

3 Can Artificial Colleagues Be Good Colleagues?

The rise of artificial intelligence systems, we noted at the outset, makes it likely that we will increasingly interact with what we call *artificial colleagues*: AI-based systems that can formulate plans and pursue actions, operate in the workplace, and

[10] Arguments for workplace democracy sometimes take a similar form, suggesting that democratising work can improve such capacities of workers (Frega et al., 2019, p. 3).

interact with workers. Drawing on our account of collegial relationships, this section offers a comprehensive assessment of the extent to which relationships with artificial colleagues can yield valuable collegial relationships goods.

Our analysis focuses on agential AI systems that operate on technology that currently exists or is foreseeable. Unlike Nyholm and Smids, we do not restrict our argument to robots. Admittedly, embodied systems are likely to be more impactful because of their greater potential to interact with humans. But in a digital workplace, unembodied AI systems can equally become colleagues. We submit that there are compelling reasons to think that such systems are limited in at least three important ways. First, they lack autonomy in a robust sense. For example, artificial colleagues cannot decide to quit their job or embark on a career change.[11] Second, they lack an inner life. By 'inner life', we mean having subjective experiences including sensations and emotions that persist over time and are self-consciously reflected.[12] Third, and plausibly given their lack of sentience and robust autonomy, such systems lack (equal) moral and political standing.[13]

Importantly, the mere fact that artificial agents have these features does not imply that relationships with them cannot provide the goods of collegial relationships. In fact, we will claim they can do so, but for the most part only to a limited degree. But this does not follow trivially: establishing it requires a close examination of how the various goods are generated.

Our discussion comes in two parts. The first sub-section addresses the individual goods of collegial relationships; the second addresses the social goods of collegial relationships. Each also briefly entertains the possibility that future AI systems overcome the stated limitations.

3.1 Artificial Colleagues and the Individual Goods of Collegial Relationships

We have distinguished three types of collegial relationships, and the (individual) goods they can bring about when successful. We now assess whether we could also obtain these goods from relationships with artificial colleagues.

[11] This is not to deny that they might be able of setting intermediate goals in narrowly specified domains (for example, figuring out the most effective way to complete a work task). But they lack the ability to independently set and develop longer-term goals in the sense of pursuing a conception of the good life. For the claim that machines are necessarily heteronomous in the Kantian sense, see Evans, Robbins, and Bryson (2023, pp. 9–10). See also Véliz's point that '[a]lgorithms […] are incapable of normatively assessing the objective for which they have been created, and modifying their behaviour accordingly' (Véliz, 2021, p. 491).

[12] For the claim that algorithms lack sentience and hence, emotions and feelings, see, e.g. Véliz (2021). For the claim that machines do not have mental states, see Johnson (2011).

[13] While there are of course deep disagreements about the grounds of personhood, it seems plausible to us that artificial systems do not currently cross the threshold. We believe that it is relatively uncontroversial that humans and artificial agents are not equal in terms of moral and political status (see, e.g., Nyholm, 2020, p. 117).

3.1.1 Professional Collegial Relationships: Solidarity and Recognition

At first sight, it might seem that we can realise the values of solidarity and recognition in professional collegial relationships with artificial colleagues. Regarding solidarity, artificial agents can often provide helpful assistance. For instance, chatbots can find and summarise information, and automated systems can detect where workers struggle and provide advice. More sophisticated systems will be able to dynamically respond to human workers, taking account of their workplace history. Regarding recognition, we can imagine artificial colleagues complimenting workers on jobs well done or sharing messages of encouragement.

We agree that professional relationships with artificial colleagues can benefit workers instrumentally, by providing useful support and feelings of validation. However, the goods of solidarity and recognition cannot be obtained to their full extent in relationships with artificial colleagues. We consider each good in turn.

Recall that the value of *solidarity* consists in the respect that we show our colleagues and in making their ends our own (Betzler & Löschke, 2021, p. 221). This form of respect involves shifting one's own perspective and sacrificing our own interests for others. We believe that the ability of artificial colleagues to do so is limited. Emotions are crucial for humans to make sense of our own experiences and of those of others: part of what it means to experience a problem as difficult is to feel frustrated or confused when confronted with it, and relieved and proud when we persevere and resolve it. We moreover draw on our own experiences to understand and appreciate others' experiences. The fact that, for all we know, artificial colleagues lack analogous emotions and sensations, limits their ability to express genuine respect.[14] Solidarity meanwhile often requires sacrificing one's own interests. Providing support to colleagues involves the cost of (temporarily) putting one's own ends and goals in the background, focusing instead on those of others.[15] But while artificial colleagues can act independently in pursuing ends, they do not set these ends themselves; their overarching goals are determined by their designers or operators.[16] Because artificial colleagues do not, then, have personal projects and interests that they pursue autonomously and care deeply about, their ability to sacrifice something of real personal value is limited.

According to Betzler and Löschke, the value of *recognition* arises in part from the fact that colleagues have an intimate knowledge of the challenges of particular tasks and can thus appreciate other workers' achievements especially well (2021, p. 222).

[14] We cannot decisively prove here that algorithms have no emotions. But we submit that the mere possibility does not invalidate our argument. We cannot prove that rocks or paintings do not have inner lives either, yet it seems plausible to assume that they do not (Véliz, 2021, p. 488). That being said, we discuss possible systems that have inner lives later in this section.

[15] In a later paper, Nyholm(2023) offers a related argument for the conclusion that robots cannot show solidarity. Nyholm rests this on the idea that offering solidary requires 'an ability to understand another person's situation […] [and] that one has certain emotional capacities' (2023, p. 12). We submit that solidarity, in addition, presupposes at the least the possibility of compromising one's personal interests.

[16] This point is also made by Evans, Robbins, and Bryson, who argue that artificial agents might have autonomy over the means by which they pursue ends, but not autonomy over the ends themselves (2023, p. 8). See also Véliz: 'algorithms are neither self-governing […] nor reasons-responsive, as no reasons could ever 'convince' them to change the goal for which they have been programmed' (Véliz, 2021, p. 492).

Artificial colleagues lack such intimate knowledge because they do not experience work in a similar way: they do not face similar physical, psychological, or emotional challenges, and tasks that are easy for humans are often difficult for machines (and vice versa).[17] Admittedly, artificial colleagues might be able to infer from statistical evidence how their human colleagues feel, and how exceptional their results are. This 'understanding' might be sufficient to offer helpful feedback to human workers. In fact, it has recently been argued that there might be no principled barriers to 'artificial empathy': artificial systems might be able to successfully emulate human empathy (Stenske & Tagesson, 2025).[18] However, we submit, any such 'understanding' must remain drastically different from human empathy grounded in first-hand experience: it remains simulated.[19] This, we submit, limits the extent to which colleagues have reason to feel appreciated in a deeper sense that presupposes genuine empathy and understanding. While artificial empathy can provide comfort and reassurance, it cannot provide recognition in the full sense.

3.1.2 Collegial Friendships: Emotional Support and Appreciation

There is a sense in which artificial colleagues may be able to offer emotional support and appreciation. As Nyholm and Smids claim, artificial colleagues can have enjoyable conversations with users (Nyholm & Smids, 2020, p. 2179) and may be designed to be self-learning such that they can build long-term relationships with humans and appear 'reliable and trustworthy' (Nyholm & Smids, 2020, p. 2180).[20] But this, we submit, is not sufficient for fully realising the collegial friendship goods of emotional support and appreciation.

When collegial friends provide emotional support and appreciation, they go beyond offering work-related advice and encouragement: they relate to us *as persons* with whom we share 'a degree of intimacy' and emotional proximity (Mlonyeni, 2023, p. 119). As we noted above, artificial colleagues cannot experience sensations and emotions in the way we do. This limits their ability to offer emotional support and appreciation. Emotional support and appreciation presuppose a commonality of experience. The knowledge that the other person experiences the world broadly as

[17] In a similar way, Nyholm argues that recognition requires an understanding that robots lack: 'to give a fellow professor, baker and so on due recognition, we need to understand and have a sense of what is involved in being a professor, baker and so on worthy of recognition' (Nyholm, 2023, p. 12).

[18] 'Artificial empathy' is a field of research in social robotics, which investigates the ability of artificial systems to emulate human empathy (Damiano et al., 2015). We take no stance on whether or when it might be desirable to implement artificial empathy (a question discussed, for example, in clinical medicine, see Sirgiovanni (2025)).

[19] One way to put this point might be that artificial colleagues can possess cognitive empathy (detect emotions), but not emotional or motivational empathy (experience emotions) (Montemayor et al., 2022, p. 1353).

[20] While Nyholm and Smids (2020, p. 2181) note that socialising with co-workers seems 'trickier' for robots to achieve, this is not our concern. Collegial friends need not socialise outside of work. In fact, the idea of *collegial* friends, as discussed by Mlonyeni (2023), is that while the relationship is friendly, it takes place in the context of the workplace. Of course, collegial friendships can grow into more general friendships, but they need not. So, even if artificial colleagues cannot socialise outside of work, *that* is not why they cannot be collegial friends.

we do and has emotional reactions to what we do and tell them partly constitutes our relationship with them, which enables the realisation of these goods.

One might object that this line of argument presupposes a too demanding a notion of collegial friendship: it requires, one might claim, an unreasonable degree of commonality. Consider, for example, John Danaher's (2019) view that friendships with robots can be mutual (based on shared values and mutual concern), authentic, and based on equal standing and diverse interaction between friends. Danaher's case relies on a behaviourist view on friendship. Standing in mutual and authentic relationships with other humans, Danaher argues, merely means 'that people engage in certain *consistent performances*' which suggest shared values (2019, p. 9).

We agree that artificial agents can consistently perform in ways that are indicative of (collegial) friendship. In this case, however, they *simulate* an inner life. *Pace* the behaviourist, we hold that merely simulating an inner life is sufficient only for *simulating* a friendship, which differs from having an actual friendship– and cannot yield the same relationship goods.[21] Note how upset we would be if we found out that our friends were actors who did not really care about us, but merely simulated their emotions– as visualised by the movie *The Truman Show*. Even if they played their parts perfectly, our anger and disappointment would be apt because we care about our friends' values and feelings: we want our friends to take a genuine interest in us. But an analogous point applies to our artificial colleagues. We have reason to think that they are incapable of having genuine feelings and emotions for us, and thus to interpret their behaviour as a simulation.[22] We have strong reason, then, to remain sceptical of their ability to actually be our collegial friends.

3.1.3 Collaborative Collegial Relationships: Shared Experience and Achievement

The potential of artificial colleagues is arguably greatest regarding collaborative collegial relationships, at least regarding the good of shared achievement.

As Nyholm and Smids (2020, p. 2179) emphasise, working with artificial colleagues can make us more effective. This might enable us to attain achievements that were previously beyond our reach. Now, one might argue that, in this case, less of the achievement is attributable to us if part of the process is automated (Danaher & Nyholm, 2021, pp. 232–34). But in so far as we are concerned with *shared* achievement, another set of considerations is central. As we have noted above, the value of shared achievement partly lies in how collaboration can itself require work and thus constitute a challenge. Succeeding in teamwork can thus increase the value of a shared achievement. On the one hand, working with machines and artificial colleagues may be thought to reduce the challenges of teamwork. For example, automated task distribution might reduce the need for coordination among humans, and artificial col-

[21] In a similar vein, Nyholm responds to Danaher's argument by arguing that 'when we value […] [our friends] as friends, we do so to a great extent because we view those attitudes and mental features as being among the reasons why they are our friends' (Nyholm, 2020, p. 118).

[22] Here, what we might call animal colleagues, such as sniffer dogs, arguably score better. There are limitations to the emotional support and appreciation that animals can offer. But they can offer some, and this is explained by the fact that their experience of life and its challenges, while radically different in many regards, is still more similar to ours.

leagues programmed to be helpful, supportive, and relatable will not provide the challenges of working with peers of varying mood, interest, and skillset. On the other hand, working with artificial colleagues can introduce new challenges. For example, to work effectively with artificial systems, workers might need to acquire additional skills. Regarding the good of shared achievement, then, we conclude that the effect of artificial colleagues could cut both ways, depending on the specific nature of the AI system and the work context.

By contrast, the experience of working together presupposes, as we noted above, a feeling of 'being in it together'. We can genuinely share experiences with human colleagues who (can) partake in our sensations, feelings, and thoughts. Artificial colleagues might be able to convincingly simulate emotional investment in shared projects, but this reintroduces the limitations identified regarding friendship: in so far as we have reason to value the shared emotional investment in a project, we have reason to think that we cannot, or only to some extent, realise the goods of shared experience and camaraderie with artificial colleagues.[23]

3.1.4 The Individual Goods of Collegiality

Many factors affect the potential of obtaining the goods of collegial relationships in relationships with artificial colleagues. While we might be able to attain similar shared achievement, we have argued that the goods of shared experience, solidarity, recognition, and emotional appreciation and support cannot fully be realised in relationships with artificial colleagues. To close, we consider two objections.

First, one might retort that, even if correct, our argument does not establish that artificial colleagues cannot be *good* colleagues, but only that they cannot be *as good* colleagues as humans. In this vein, Nyholm (2024) adopts a suggestion from Ryland (2021), who argues that friendship comes in degrees (roughly, we can see someone as more or less of a friend). Similarly, Nyholm argues, someone can be a better or worse colleague: 'robots could be good colleagues to a certain degree, and perhaps only a very limited degree' (Nyholm, 2024, p. 115). This suggestion is compatible with our argument. Colleagues can differ in their abilities to show solidarity, or emotional support, or obtain shared achievements. However, if our argument is correct, then it shows something important: in contrast to human colleagues who might vary in their individual abilities, resources, and motivations, the noted features of artificial colleagues place hard limits on the extent to which they can realise the goods of collegial relationships– or, in short, the extent to which they can be good colleagues.

Second, one might object that our arguments are contingent: they would not hold for advanced AI systems that might one day exhibit robust autonomy and inner lives. Let us, for the sake of argument, entertain the possibility of such systems (though we believe that it is difficult to concretely picture this). For all we have said, such advanced artificial colleagues could indeed experience genuine empathy, truly sac-

[23] One might respond, once more, that perhaps the mere perception that the artificial colleagues share our experiences could, even if illusory, be sufficient. In this case, whether humans can experience camaraderie with artificial colleagues would turn into an empirical question. So far, studies show that humans find working with other humans more meaningful and motivating (see, e.g., Sadeghian & Hassenzahl, 2022).

rifice their interests, and be emotionally invested in their human colleagues and the success of shared projects. Even so, we believe that there is good reason to resist the conclusion that they could fully realise all collegial relationship goods. Yes, such advanced artificial agents would have inner lives, but these would presumably be very different from our own, and the parameters within which they would set their ends would likely be very different, too. Such artificial agents might have perceptive capabilities that we do not have (they might have sensors that detect radiation) or lack ones we have (they might not be able to taste or smell). They might not be embodied, or they might be instantiated in multiple bodies at once. Their attention might not be limited in the way ours is. Their fundamental needs and wants might consequently differ radically from ours. We submit that these differences will likely limit the extent to which we could have shared experiences and reciprocal relationships. So, while it is true that our main argument is restricted to AI systems that have certain limitations, we believe that there is good reason to think that even more advanced artificial colleagues, should they one day exist, would be unable to fully realise all collegial relationship goods.

3.2 Artificial Colleagues and the Social Goods of Collegial Relationships

We now turn to the social goods of collegiality: empowerment and social cohesion. We will argue that while artificial colleagues can support the realisation of these goods, they cannot fully realise them. This is because how these goods can be realised in relationships with artificial colleagues differs in morally significant ways from the ways in which they can be generated in relationships between human workers.

To set out our argument, consider a line of reasoning that suggests that widespread relationships with artificial colleagues can give rise to empowerment and social cohesion. Suppose that artificial agents can be designed as instructors, which are capable of reliably sensing a person's views, moods, and dispositions, and predicting how they will respond to specific inputs. Moreover, they are capable of adopting a variety of viewpoints and assuming the role of other agents. By various methods, such as reasoning through hypothetical and actual situations, these systems can teach humans a variety of social skills, including, e.g., anger management, negotiation strategies, and conflict resolution.[24] Now suppose that our artificial colleagues could be programmed in this way, and that they could thereby, in addition to working in particular professions, assume the capacity of teaching us social skills. Could such systems not help foster empowerment, in the above-defined sense of practising skills and competences critical to our ability to act as democratic citizens, and social cohesion, by instructing us in ways that lead to a reduction in conflict and greater cooperation?

We believe that even in the case of such highly capable artificial colleagues, there are two reasons to be sceptical of their ability to realise the social goods of collegial relationships.

The first reason is that interacting with human colleagues directly connects us to other members of society who are our moral and political equals, whereas interacting

[24] While these systems are more advanced than currently existing technology, they fit within our paradigm: their ability to sense our inner lives does not imply that they have an inner life of their own.

with artificial colleagues does not. Interacting with artificial colleagues would be, at best, *training* for social relationships with other humans, rather than an instance of *practising* such relationships (as in interactions with human colleagues). So, while relationships with artificial colleagues could generate empowerment and social cohesion, they would do so through a different mechanism. The goods would be generated indirectly, through training that influences our relationships with other humans, rather than directly, through interacting with peers. Interacting with artificial colleagues does not, then, constitute a practice of social relationships among equals in the same way.

The second reason is that even where our interactions with these artificial colleagues are conducive to empowerment and cohesion, they are not, in the primary sense, interactions *within collegial relationships*. As the example indicates, artificial colleagues would assume the roles of teachers or trainers and, in this respect, a decidedly asymmetrical position. In fact, these artificial colleagues would have been designed specifically to have the educational effect of boosting the social goods of empowerment and cohesion. It is true that they might still be our colleagues: we might still interact with them at work. But any collegial relationship is here merely incidental: the artificial systems could equally have this effect if we interacted with them in other contexts, such as a classroom or a therapy session. In the case of successful collegial relationships with humans, by contrast, empowerment and social cohesion are an unintended consequence of the free interactions among colleagues.

The centralised, external control that AI systems can facilitate poses an additional potential threat to empowerment and social cohesion. Because AI systems are prohibitively expensive to develop, but often have near zero marginal costs of provision, there is a strong economic rationale for assuming that at least within firms, control over the AI systems will be centralised– in the hands of management. There is thus good reason to think that, as Evans, Robbins, and Bryson (2023, 13) argue, the rise of such AI systems will shift the balance of power in favour of owners of capital.[25] This centralisation of control, in combination with the ability of managers to scour large data sets, directly undercuts empowerment of ordinary citizens; in addition, it can also constitute a threat to social cohesion, at least where AI colleagues remain inscrutable and opaque, and their presence sows distrust.

In sum, we believe that relationships with artificial colleagues can only generate the social goods of collegial relationships to a limited degree. Even if we could find a way to substitute for the individual goods of collegial relationships– be it outside of work or inside– a loss in human collegial relationships would thus come at an important social cost. Moreover, we have identified opportunities and risks even in cases where artificial colleagues complement, but do not entirely replace, human colleagues: depending on how they are designed and employed, they might support training and development, but also risk exacerbating power imbalances between workers and management.

[25] Evans, Robbins, and Bryson (2023, pp. 13–14) argue that talk about 'collaboration' with AI is misplaced because real collaboration involves joint planning and decision-making, which AI systems are incapable of, since they act as 'surrogate agents' for those employing the systems and do not set their own ends.

Again, one might object that our argument fails to apply to advanced artificial colleagues, should they one day enjoy equal moral and political standing. Could relationships with such advanced artificial colleagues equally generate social cohesion and empowerment? We submit that this is difficult to ascertain because it is unclear what exactly our political community would then look like. But we believe that, at a minimum, it would not evidently follow that relationships with such artificial colleagues could generate these goods. Artificial agents might be so alien that we would struggle to understand them, in which case even frequent interactions might fail to generate social cohesion. And if their form of life necessitates divergent forms of political organisation or organisation, then even frequent interactions with artificial colleagues might fail to empower us for this political context. There is thus reason to think that even advanced artificial colleagues would be limited in their ability to generate the social goods of collegial relationships.

4 The Spectre of Artificial Colleagues: Policy Implications

In this final section, we briefly ask: if our account of the value of collegial relationships and our assessment of the limits of artificial colleagues is correct, what are the policy implications? Given the diversity of professional settings, and the variety of artificial colleagues imaginable, our discussion must remain cursory. But we believe that our assessment of the limitations of artificial colleagues nonetheless allows us to sharply capture some insights; and that it supports what we call the *Principle of Strict Priority of Human Collegial Relationships.*

The first and perhaps rather obvious insight is that if our account of the value of collegial relationships and of the limits of artificial colleagues is correct, then artificial colleagues will be unable to fully replace humans in providing collegial relationship goods. To be clear, our argument does not imply that agential AI systems could not be great team members, or co-workers, or live up to the relational ideal of some another, perhaps yet-to-be-conceptually-engineered relationship. But if we are right that there is a specific value to collegial relationships, then the inability of agential AI systems to realise the goods of collegial relationships remains an important limitation. It indicates that the predicted rise of artificial colleagues does not diminish the relevance of maintaining work conditions conducive to successful collegial relationships *among humans*.

The second, related insight is that we should remain attentive to the risk that artificial colleagues crowd out valuable collegial relationships among humans. If our argument concerning their limitations holds, then the introduction of artificial colleagues that replace human ones is, *ceteris paribus*, not a welcome result. This is not to say that we should favour human colleagues at all costs. For one thing, human collegial relationships can also be a source of bads: colleagues can be biased; they can bully others. Moreover, other benefits, like efficiency gains, might tilt the balance in favour of artificial colleagues. But interactions with artificial colleagues cannot replace the *opportunity* to have successful human collegial relationships, so it matters if a push towards artificial colleagues crowds out this opportunity.

The third insight is that the presence of artificial colleagues can affect relationships among human colleagues. The effects need not be negative, but we submit that there is a risk that they are. Two scenarios illustrate this possibility. First, consider a *malign scenario* where artificial colleagues are centrally operated by managers (or capital holders) to continuously monitor human workers.[26] Suppose artificial colleagues are trained to infer insights about human workers from their observation of interactions among humans. Here, the presence of artificial colleagues can undermine human collegial relationships– by diminishing trust and by intruding on their privacy. Humans might have reason, for fear of repercussions, to refrain from ordinary collegial behaviour, such as taking a break to chat or bonding over a rant against workplace conditions.[27] Next, consider a *benign scenario*, where the sole purpose of artificial colleagues is to display the most virtuous behaviours one would hope to elicit from colleagues (suppose they are controlled by the workers collectively). Even these artificial colleagues could undermine human collegial relationships, albeit in a more indirect way. For example, by setting an unattainable standard, they could cast a shadow over humans who generally fail to live up to it. This could bring about a form of alienation: humans might, by contrast, be seen as impulsive, emotional, and generally bothersome. This could bring about a withdrawal from human collegial relationships or a reassessment of the behaviour of colleagues. While the detailed effects of having artificial colleagues would likely be more complex in practice, these two extreme scenarios illustrate that artificial colleagues could drastically alter relationships *between* human colleagues.

What are the policy implications of these findings? We believe that our discussion of the limitations of artificial colleagues and their potential interference with valuable collegial relationships among humans motivates the following *governance principle*:

> Principle of Strict Priority of Human Collegial Relationships In evaluating the effects of policy proposals *on collegial relationships*, strict priority must be assigned to their effects on the opportunity to realise valuable human collegial relationships.

We propose the *Principle of Strict Priority of Human Collegial Relationships* as a governance principle.[28] It is not meant as an ethical principle: we do not claim that every individual, such as a manager, is ethically obligated to give priority to human collegial relationships (in evaluating the effects of policy proposals on collegial relationships). Neither is it meant as a legal principle: we do not propose that any individual is, or should be, legally required to so weigh human collegial relationships. In speaking of it as a governance principle, we instead take it to be a principle that should guide policy making. To make this clearer, consider the case of a tragedy of

[26] This resembles the scenario envisioned by Evans, Robbins, and Bryson, where '[c]apital then becomes an arbitrary number of Potemkin co-workers masquerading as labor, presumably defusing true labor's capacities to organize and even self-represent' (2023, p. 13).

[27] Beyond surveillance, they might be tasked with surreptitiously influencing workers, which might interfere with relationships among humans in more complex and covert ways.

[28] We are grateful to an anonymous reviewer for noting these different possible interpretations, and for pressing us to be clearer on the status and content of this principle.

the commons (Hardin, 1968). One could propose an ethical or a legal principle here, which stipulates how one is morally or legally required to act (e.g., do not use more than a sustainable fair share of the commons). By contrast, a governance principle would stipulate how policy makers should seek to regulate the use of the commons (e.g., create enforceable rules that prevent depletion). While the principle we propose here is less determinate– it does not directly stipulate which action to take, but rather how to weigh certain considerations– it, too, concerns policy makers.

But policy making can take at least two forms: self-regulation within an industry and state regulation. The governance principle is thus meant to apply to both industry leaders (in their capacity as joint rule-setters of the industry), and to state policy makers (in their role as regulators of the workplace). In either case, the principle makes a fairly narrow demand. It demands that, in so far as policy makers are concerned with the effects of artificial colleagues on collegial relationships (and the principle only makes a claim about this), they should give priority to the effects on human collegial relationships. Note that this in no way implies that the introduction of artificial colleagues should be prohibited. In fact, the principle does not even commit to the view that the effect of artificial colleagues on collegial relationships will be negative. In only claims that, *in so far as we are concerned with the effects of policy proposals on collegial relationships* (which will generally be one consideration among many), we should give priority to relationships among human colleagues (or, more specifically, to effects on the opportunity to form such relationships).

To illustrate, consider the question whether, and if so, in what form, warehouses should be permitted or incentivised to introduce humanoid robots with advanced sensor and speech capacities to work alongside humans (be it in form of government regulation or industry self-regulation). A wide range of considerations are relevant here. All the principle demands in this case is that, in so far as the effects on collegial relationships are concerned, one should focus on how the policy will impact the opportunity of workers to form human collegial relationships. This implies that an enhanced opportunity to form relationships with artificial colleagues cannot outweigh a deteriorated opportunity to form collegial relationships with humans. But the principle does not yield any verdict on whether the introduction of such robots should be permitted or incentivised. This is contingent on a host of other issues. What is the net effect of introducing artificial colleagues on the opportunity to have human collegial relationships? How should we balance the effect of the policy on collegial relationships against its other effects? Finally, how should we weigh the sum of these effects against the freedom of employers to control their workplaces?[29]

In one sense, the proposed principle is modest: it does not directly prescribe any specific policy. In another sense, however, we believe it has significant implications for the justifiability of policies. As we move towards a world where artificial intelligence systems assume an increasing range of functions, they will enter many work-

[29] For the principle to have bearing on *questions of permission*, one needs to make the further assumption that some restrictions on the (negative) freedom of employers or capital owners to design their workplace as they see fit are justifiable. While radical libertarians might deny this assumption, we believe it is widely accepted. In practice, all employers confront some rules and restrictions. And philosophers have argued that even if freedom were our only concern, we would confront a trade-off between the negative freedom of employers and the positive and republican freedom of workers (see, e.g., Anderson, 2017).

places. In demanding that, in so far as collegial relationships are concerned, policy makers ought to give strict priority to the opportunity to form relationships among humans, the principle rules out certain trade-offs. For example, it bars balancing a loss in (opportunity to realise) human collegial relationships with enhanced (opportunity to realise) relationships with artificial colleagues. This precludes one important potential argument by proponents of artificial colleagues, namely that they can fully replace humans in their function as colleagues. And it orients the focus of the debate, in so far as collegial relationships are concerned, clearly towards the impact of policies on human collegial relationships.

5 Conclusion

In this paper, we have offered an assessment of the extent to which relationships with artificial colleagues can provide workers with the goods of collegial relationships and discussed its practical implications. We did so in three steps. We first set out a comprehensive account of the goods of collegial relationships, which contributes to the existing literature by introducing a new distinction between individual and social goods of collegial relationships. We then employed this account to assess to what extent artificial colleagues can realise the various collegial relationship goods. Our conclusions, while mixed, are overall sceptical. The extent to which relationships with artificial colleagues can provide workers with the individual goods is limited in important ways (although we do not deny that such relationships can be beneficial for workers); and they can at best support the realisation of the social goods. In turning to practical implications, we noted the upshot that we should be attentive to how artificial colleagues affect opportunities to form collegial relationships among humans – directly, through crowding out, or indirectly, through interference. And we proposed a governance principle to guide policy makers in the evaluation of policies regarding artificial colleagues. This principle demands giving strict priority, in so far as their effects on collegial relationships are concerned, to the impact of policies on opportunities to have collegial relationships among humans.

Acknowledgements The authors thank the audience of the workshop on partiality and relationships with AI at Ludwig Maximilian University Munich for helpful discussion. We are especially grateful to Sven Nyholm for helpful comments and suggestions, and to Tom Douglas for helpful written comments.

Author Contributions Both authors have contributed equally to the conception, drafting, and revising of the manuscript. Authors are listed in alphabetical order.

Funding Friedemann Bieber's work on this article was supported through the University of Zurich's Research Priority Program "URPP Equality of Opportunity".

Declarations

Ethical Approval N/A.

Consent To Participate N/A.

Consent To Publish N/A.

Competing Interests The authors declare they have no financial interests.

References

Althorpe, C. (2022). Meaningful work, nonperfectionism, and reciprocity. *Critical Review of International Social and Political Philosophy*, 1–22. https://doi.org/10.1080/13698230.2022.2137753

Altman, S. (2025). *Reflections.* Retrieved March 27, 2025, from https://blog.samaltman.com/reflections

Amodei, D. (2024). *Machines of Loving Grace: How AI Could Transform the World for the Better.* Retrieved March 27, 2025, from https://darioamodei.com/machines-of-loving-grace

Anderson, E. (2017). *Private government: How employers rule our lives (and why we don't talk about It)*. Princeton University Press.

BBC (2024). *We'll need universal basic income - AI 'godfather'*. BBC News. Retrieved March 27, 2025, from https://www.bbc.com/news/articles/cnd607ekl99o

Betzler, M., & Löschke, J. (2021). Collegial relationships. *Ethical Theory and Moral Practice*, *24*(1), 213–229. https://doi.org/10.1007/s10677-021-10165-9

Bieber, F., & Unruh, C. (forthcoming). Collegial Relationships and the Non-Monetary Goods of Work. Philosophers' Imprint. https://doi.org/10.3998/phimp.4828

Bradford, G. (2015). *Achievement*. Oxford University Press.

Brighouse, H., & Swift, A. (2014). *Family values: The ethics of Parent-Child relationships*. Princeton University Press.

Damiano, L., Dumouchel, P., & Lehmann, H. (2015). Artificial empathy: An interdisciplinary investigation. *International Journal of Social Robotics*, *7*(1), 3–5. https://doi.org/10.1007/s12369-014-0259-6

Danaher, J. (2019). The philosophical case for robot friendship. *Journal of Posthuman Studies*, *3*(1), 5–24. https://doi.org/10.5325/jpoststud.3.1.0005

Danaher, J., & Nyholm, S. (2021). Automation, work and the achievement gap. *AI and Ethics*, *1*(3), 227–237. https://doi.org/10.1007/s43681-020-00028-x

Estlund, C. (2003). *Working together: How workplace bonds strengthen a diverse democracy*. Oxford University Press.

Evans, K. D., Robbins, S. A., & Bryson, J. J. (2023). Do we collaborate with what we design? *Topics in Cognitive Science*, 1–20. https://doi.org/10.1111/tops.12682

Fleming, M., Li, W., & Thompson, N. C. (2024). *The Last Mile Problem in AI: Why job automation will be slower than technological progress suggests.* Brookings Institute. Retrieved March 27, 2025, from https://www.brookings.edu/articles/the-last-mile-problem-in-ai/

Frega, R., Herzog, L., & Neuhäuser, C. (2019). Workplace Democracy—The recent debate. *Philosophy Compass*, *14*(4), e12574. https://doi.org/10.1111/phc3.12574

Friedman, M. (1989). Friendship and moral growth. *The Journal of Value Inquiry*, *23*(1), 3–13. https://doi.org/10.1007/BF00138682

Gabriel, I., Manzini, A., Keeling, G., Hendricks, L. A., Rieser, V., Iqbal, H., Tomašev, N., Ktena, I., Kenton, Z., Rodriguez, M., El-Sayed, S., Brown, S., Akbulut, C., Trask, A., Hughes, E., Bergman, A. S., Shelby, R., Marchal, N., Griffin, C., & Manyika, J. (2024). The Ethics of Advanced AI Assistants. arXiv. Retrieved March 27, 2025, from https://arxiv.org/abs/2404.16244.

Gheaus, A. (2022). Personal Relationship Goods. In Zalta, E.N. & Nodelmann, U. (Eds.), *Stanford Encyclopedia of Philosophy*, Winter 2022 Edition. https://plato.stanford.edu/archives/win2022/entries/personal-relationship-goods/

Hardin, G. (1968). The tragedy of the commons. *Science*, *162*(3859), 1243–1248. https://doi.org/10.1126/science.162.3859.1243

Helm, B. (2023). Friendship. In Zalta, E.N. & Nodelmann, U. (Eds.), *Stanford Encyclopedia of Philosophy*, Fall 2023 Edition. https://plato.stanford.edu/archives/fall2023/entries/friendship/

Johnson, D. G. (2011). Computer systems: moral entities but not moral agents. In M. Anderson, & S. L. Anderson (Eds.), *Machine ethics* (pp. 168–183). Cambridge University Press.

Lintott, S. (2015). Friendship and bias: Ethical and epistemic considerations. *Journal of Social Philosophy*, *46*(3), 318–339. https://doi.org/10.1111/josp.12105

Mlonyeni, P. M. T. (2023). Collegiality, friendship, and the value of remote work. *Ethical Theory and Moral Practice*, *26*(1), 113–126. https://doi.org/10.1007/s10677-023-10368-2

Montemayor, C., Halpern, J., & Fairweather, A. (2022). In principle Obstacles for empathic AI: Why we can't replace human empathy in healthcare. *AI & SOCIETY*, *37*(4), 1353–1359. https://doi.org/10.1007/s00146-021-01230-z

Nyholm, S. (2020). *Humans and robots: Ethics, agency, and anthropomorphism*. Rowman & Littlefield.

Nyholm, S. (2023). Can a robot be a (Good) colleague?? In M. Filimowicz (Ed.), *AI and the future of creative work* (pp. 1–17). Routledge.

Nyholm, S. (2024). AI, robot Co-Workers and humans. In A. Elliott (Ed.), *The de Gruyter handbook of artificial intelligence, identity and technology studies* (pp. 101–120). De Gruyter.

Nyholm, S., & Smids, J. (2020). Can a robot be a good colleague?? *Science and Engineering Ethics*, *26*(4), 2169–2188. https://doi.org/10.1007/s11948-019-00172-6

Roessler, B. (2012). Meaningful work: Arguments from autonomy. *Journal of Political Philosophy*, *20*(1), 71–93. https://doi.org/10.1111/j.1467-9760.2011.00408.x

Ryland, H. (2021). It's friendship, Jim, but not as we know it: A Degrees-of-Friendship view of Human-Robot friendships. *Minds and Machines*, *31*(3), 377–393. https://doi.org/10.1007/s11023-021-09560-z

Sadeghian, S., & Hassenzahl, M. (2022). The Artificial Colleague: Evaluation of Work Satisfaction in Collaboration with Non-Human Coworkers. *27th International Conference on Intelligent User Interfaces*, 27–35. Helsinki Finland: ACM. https://doi.org/10.1145/3490099.3511128

Schwartz, A. (1982). Meaningful work. *Ethics*, *92*(4), 634–646. https://doi.org/10.1086/292380

Seglow, J. (2013). *Defending associative duties*. Routledge.

Sirgiovanni, E. (2025). Should Doctor robot possess moral empathy?? *Bioethics*, *39*(1), 98–107. https://doi.org/10.1111/bioe.13345

Stenske, J., & Tagesson, A. (2025). The prospects of artificial empathy: A question of attitude?? *Frontiers in Artificial Intelligence and Applications*, *397*, 149–156.

Susskind, D. (2020). *A world without work*. Allen Lane.

Telfer, E. (1971). Friendship. *Proceedings of the Aristotelian Society*, *71*(1), 223–242. https://doi.org/10.1093/aristotelian/71.1.223

Thomas, L. (1987). Friendship. *Synthese*, Kurt Baier Festschrift, Part II, *72*(2), 217–36.

Thomas, L. (1999). Split-Level equality: Mixing love and equality. In S. Babbit, & S. Campbell (Eds.), *Racism and philosophy* (pp. 189–201). Cornell University Press.

Véliz, C. (2021). Moral zombies: Why algorithms are not moral agents. *AI & SOCIETY*, *36*(2), 487–497. https://doi.org/10.1007/s00146-021-01189-x

Velleman, J. D. (1999). Love as a moral emotion. *Ethics*, *109*(2), 338–374. https://doi.org/10.1086/233898