EDITORIAL

Miriam Meckel
is a Professor for Communication Management at the University of St. Gallen in Switzerland. A journalist and former State Secretary and government spokeswomen, she is also the Co-Founder and CEO of ada, a learning platform for innovative corporate training and development.

Léa Steinacker
is a researcher at the University of St. Gallen in Switzerland with a research focus on the socio-technological implications of artificial intelligence. A journalist and practitioner of innovation in digital media, she is the Co-Founder and COO of ada.

Christian Fieseler is Professor for Communication Management at BI Norwegian Business School and a director of the Nordic Centre for Internet and Society.

In 1842, Ada Lovelace first mused about "a new, a vast and powerful language (…) developed for the future use of analysis" (p. 23) as she formulated the first algorithmic instructions to be coded into a machine. It was a ground-breaking vision for the possibilities of computing. A century later, Alan Turing, in his landmark 1950 article Computing Machinery and Intelligence, stipulated that any calculation which can be performed by a smart human mathematician could also be performed by a machine or, as he was later paraphrased by inventor Daniel Hillis, by a "stupid but meticulous clerk" following a "simple set of rules" (1999, p. 63). Over the past decades, various disciplines have striven to dream up, design, and construct such a clerk. And with astounding success. Since the visions of Lovelace and Turing, advances in computing machinery and algorithms have effectively ushered in an age of ever more intelligent machines. Our 'stupid but meticulous clerks' have, in many ways, become much less stupid as they increasingly build an important backbone of our economies, societies and culture. As intelligent machines populate our everyday lives, facilitate decision-making, and contribute to artistic expression, many questions remain: What is the delineation between artificial intelligence and automated stupidity? What role do humans play in constructing both? How are we to think of intelligent machines as moral agents? Can they be held accountable? What are meaningful boundaries between humans and machines?

With Morals & Machines, we want to provide a forum for informed debate around which type of social contract we want to construct for living well with these increasingly capable machines. Set against the backdrop of ubiquitous applications of autonomous systems, Morals & Machines provides a space for a curious community seeking to explore the intersections and dissolving boundaries between moral reasoning and technological creation, between machine behavior and social implications, between conceptual frameworks, regulations, and practical implementation of technologies. This journal will be based on two overarching guideposts: the agency of machines, and the morals we want to instill them with, if any.

SIRI, SMART-HOMES AND SELF-DRIVING CARS: SOPHISTICATED TOOLS OR SOCIAL AGENTS?

Our first overarching concern with this journal is the question of what kind of agency can be assigned to machines. This question can be answered from many different disciplinary backgrounds in a variety of ways. A few demarcations seem to be however of utter importance in this field. Firstly, machines differ in the degree to which they can be assigned agency. Rammert (2008) argues that they range from passive tools to self-coordinating co-operative systems. A hammer, for example, is completely driven from outside. Semi-active machines display some aspects of self-acting, such as a record player. Re-active or cybernetic systems entail feedback loops, such as a thermostat-driven climate control. Pro-active systems are based on self-activating and amplifying programs, such as AI-based voice assistants. Finally, there are distributed and self-coordinating co-operative systems, such as smart homes or self-driving cars.

Morals & Machines is particularly interested in autonomous and agentic machines, self-learning and self-replicating machines, connective machines, creative, assertive, and curious machines, and embodied and virtual machines. For all these kinds of machines, we can adapt Marshall McLuhan's (1964) famous description of the medium changing the nature of the conveyed content ("the medium is the message"). Each kind of machine continually changes the moral message towards the users, collaborators or co-agents as well as the moral systems upon which their human conceptions and interactions are built.

Every type of potential machine agency will be based on the visions of those who have conceived of the technology, inscribed in its algorithmic models and its rules for autonomy. In their essay on machine behavior, Rahwan et al. (2019) suggest focusing on three challenges related to these sociotechnical systems: The "ubiquity of algorithms", the "complexity and opacity of algorithms", and the "algorithms' beneficial and detrimental effect on humanity" (p. 478). This journal aims to analyze, interpret, and discuss the respective sociotechnical systems and moral frameworks by which autonomous machines are powered as well as the impact that these frameworks can have on human machine interaction and collaboration. Rahwan et al. (2019) further differentiate between "individual machine behavior", "collective machine behavior", and "hybrid human-machine behavior" (p. 481). While we are particularly interested in the latter, contributions to this journal will cover all three manifestations of a wide range of machine agency.

HUMAN-CENTRIC DESIGN: SAFE, MORAL AND TRUSTWORTHY MACHINES?

Our second overarching concern is with the kinds of morality we want to instill into these machines as each system is constructed by inherent normative forces. In the debate about the moral impacts of autonomous machines, current discussions often focus on fairness, transparency and accountability, exposing biased and problematic AI initiatives. In many instances, problems arise not out of malice but out of ignorance, a deficit of knowledge or context, or inadequate tools. This debate is also a political one, around which social contract we want to abide by when integrating machines into society, and how these choices affect individuals. In April of this year, for instance, the EU Commission (2021) published the proposal for a regulation laying down harmonized rules on artificial intelligence, the first of this scope worldwide, which has since evoked several reactions. In the proposal, the EU argues that "Rules for AI […] should […] be human centric, so that people can trust that the technology is used in a way that is safe and compliant with the law, including the respect of fundamental rights" (p. 1). Critics challenge this approach by the assumed fact that artificial intelligence does not lend itself to easily drawn boundaries, simply due to the nature of the technology. In Atlas of AI, Kate Crawford calls for considering the technology beyond the technical breakthroughs, urging that "to understand how AI is fundamentally political, we need to go beyond neural nets and statistical pattern recognition to instead ask what is being optimized, and for whom, and who gets to decide. Then we can trace the implications of those choices" (Crawford, 2021, p.9). This socially embedded nature of the technology, be it AI, the blockchain or upcoming quantum computing, must be described and analyzed, particularly regarding the impact it will have on human-machine-collaboration, during this ongoing shift of authority and accountability.

MORALS AND MACHINES: BOLD HUMAN THINKERS WANTED

Against these two broad concerns, with Morals & Machines we want to expand the understanding of societal impacts of machines by enhancing knowledge sharing and coordination, and creating new knowledge pathways for scholars, industry, and policymakers from around the world. Melding the aspects of morals and machines, the journal aims to find pathways for pressing issues such as algorithmic accountability, risk assessments, governmental automated decision-making, bias and discrimination, privacy protections, and efforts at negotiating the contours of humanness. Our journal will encourage critical debate around these themes and aims to enhance the capabilities of key change agents who can most contribute to and benefit from additional work on the societal impacts of artificial intelligence (AI).

We especially encourage bold, controversial, and explorative conceptual contributions that propose new, speculative, and unexpected solutions to imbue machines with morality, or critique the lack thereof. The journal is open to broad conceptualizations of machines, in particular any forms of autonomous and agentic entities, of self-learning and self-replicating mechanisms, of connective sensors, as well as of embodied and virtual avatars. Likewise, the journal is interested in any well-reasoned moral lens on machines and managing machine behavior, ranging from regulatory, to philosophical, to lay, to religious, to non-moral, and beyond. We hope to present ideas that shape law and regulation, as well as managerial, social, and cultural action around these issues and develop working compromises for messy dilemmas. We welcome manuscripts at the formative stages of thinking, that ideally contribute towards building new theory, methodology, and novel ways of organizing and governing. We are especially interested in contributions that go beyond analysis and provide implications for conceivable courses of action.

We thought it fitting to start the journal with the theme of hybridity, as it speaks to the interconnected character of morals that mediated by socially embedded machines. In a hybrid, two forms join to create an entity that is neither one nor the other; so three forms are involved or implied. Being human in an age of ever more intelligent machines entails an ever-increasing reliance on and entanglement with nonhuman materiality, such

as 'smart' connected ecosystems, wearable devices, and advanced robotics, making ours a progressively hybrid existence (Soekadar et al., this issue). This phenomenon has been theorized in terms of the cyborg (Haraway, 1991; Zylinska, 2002), or the posthuman (Hayles, 2008). Hybridity and Hybrid Evolution might be perceived as a threat to our human exceptionalism: if I can be part machine and part human, then the human part does not look so exceptional after all. But hybridity can also be an acknowledgment that all living beings exist on a continuum; that nothing about any of us can ever be "pure" (cf. Lewis, 2020).

In this first issue, the theme of hybridity underpins our eight contributions. Together they demonstrate that there is an interesting landscape ranging from machines shaping our perception of reality, to those shaping our behaviors, and those expanding our capabilities. These are reflected in our first four contributions, followed by four articles that address potential regulatory and self-regulatory reactions to this new landscape.

The issue is introduced by Miriam Meckel and Léa Steinacker's reflections on the increasing impact of deepfakes, imagery manipulated and created through AI. Painting a picture of a foreseeable hybrid reality, where technologically driven fabrications may develop a societal reality of their own, they caution against a possibly distorted marketplace of truths. To safeguard against this, they argue, we have to agree on new technological, deliberative and political measures.

Next, Thales Bertaglia, Adrien Dubois, and Catalina Goanta share their observations of how the largely automated reward system for social media content creators incentivizes the creation of controversy to boost visibility. In their discussion of this practice, they not only highlight the contradictions present in many platform governance systems but also describe the fine balancing act creators discover for themselves between chasing clout, hedging their bets against automated content moderation, and possible repercussions.

Surjo Soekadar, Jennifer Chandler, Marcello Ienca, and Christoph Bublitz write about the implications of the hybrid mind, an assemblage between the human and technology. In their article they raise pressing questions regarding our perceptions of self, and blurring boundaries between our agency and those of the technology that might increasingly co-construct our minds.

Shohini Ghose provides us with an essay on the implications of quantum computing for our social and political systems. Relating to us postulates from quantum science such as the notion of entanglement and superposition, her article provides an inspiration for how to approach the looming social implications of this new computing paradigm in a more fluid and inclusive mindset.

tValentin Jeutner, in his article, reflects on the impact of such a quantum future on the protection of rights and power relations. Pointing to both the implications of designing and operating quantum computers, he proposes a new 'quantum imperative', cautioning against creating or exacerbating inequalities, and showing regulatory pathways to avoid undermining individual autonomy, and provide consultation for those affected.

Henrik Skaug Sætra and Eduard Fosch-Villaronga explore to what degree we should or should not restrict the development of AI. Arguing along four conjectures, they conclude against an ex-ante regulation of science and placing the burden of ethical assessment solely on innovators. Instead, they call for ethicists and politicians to step up more effectively in evaluating and regulating the science produced, so that it is not uncritically applied in society.

Alexander Buhmann and Christian Fieseler embed the current debate surrounding the opaque tendencies of autonomous technologies in the wider discourse on the design of responsible innovation. Arguing against solely pragmatic approaches to create legitimacy among stakeholders for technologies they largely have to trust, such as the provision of engagement spaces or the quest for demographic inclusivity, they propose a number of communicative principles. True legitimacy and acceptance rests, they argue, also on the discursive quality of how technological principles are agreed upon, not solely the provision of forums.

Finally, Sofia Ranchordas argues against the ill-conceived notion that regulation is necessarily diminishing innovation, and instead presents new approaches in the legal environment to use experimental regulation and sandboxes to curate emerging new technology. Straddling the line between the need to prevent harm and to help create a better understanding of regulatory impact, her article discusses the learnings from recent efforts and proposes procedures to create the largest benefits of such new regulatory tools.

Our gratitude goes to Thomas Gottlöber for his impetus in late 2020 that we launch this journal. Together with the journal's editorial officer, Sandra Frey, we have much enjoyed taking the initial idea from Nomos on an interdisciplinary journey of exploration, in the spirit of Ada Lovelace, who reportedly said: "I never am really satisfied that I understand anything; because, understand it well as I may, my comprehension can only be an infinitesimal fraction of all I want to understand about the many connections and relations which occur to me" (Ada Lovelace Institute, 2019).

REFERENCES:

Crawford, K. (2021). The Atlas of AI. Yale University Press.

European Commission (2021). Proposal for a Regulation of the European Parliament and of the council laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. Brussels.

Haraway, D. (1991). An ironic dream of a common language for women in the integrated circuit. Philosophy of Technology. Blackwell Publishing Ltd.

Hayles, N. K. (2008). How we became posthuman: Virtual bodies in cybernetics, literature, and informatics. University of Chicago Press.

Hillis, W. D. (1999). The pattern on the stone: the simple ideas that make computers work. Basic Books.

Lewis, S. A. (2020). Hybridity and the Cyborg. International Encyclopedia of Human Geography, 7, 129 – 135.

Ada Lovelace Institute (2019). Celebrating Ada Lovelace Day: what Ada means to us. Available online at: https://www.adalovelaceinstitute.org/blog/celebrating-ada-lovelace-day/

McLuhan, M. (1964). Understanding media. Signet.

Menabrea, L. F., & Lovelace, A. (1842 / 1843). Sketch of the analytical engine invented by Charles Babbage. R. & J. E. Taylor

Rahwan, I. et al. (2019). Machine behaviour. Nature 568: 477-486. 10.1038/s41586-019-1138-y.

Rammert, W. (2008). Where the action is: Distributed agency between humans, machines, and programs. In U. Seifert, J. H. Kim & A. Moore (Eds.), Paradoxes of interactivity. Bielefeld, Germany: Transcript, 62–91.

Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433.

Zylinska, J. (Ed.). (2002). The cyborg experiments: The extensions of the body in the media age. A&C Black.