Editors' Preface

The idea for this handbook arose in late 2017, with the working title *Handbook of Ethics of AI in Context*. By the time solicitations went out to potential contributors in the summer of 2018, its title had been streamlined to *Handbook of Ethics of AI*. Its essentially contextual approach, however, remained unchanged: it is a broadly conceived and framed interdisciplinary and international collection, designed to capture and shape much needed reflection on normative frameworks for the production, application, and use of artificial intelligence in diverse spheres of individual, commercial, social, and public life.

The approach to the ethics of AI that runs through this handbook is contextual in four senses:

- it locates ethical analysis of artificial intelligence in the context of other modes of normative analysis, including legal, regulatory, philosophical, and policy approaches,
- it interrogates artificial intelligence within the context of related modes of technological innovation, including machine learning, Big Data, and robotics,
- it is interdisciplinary from the ground up, broadening the conversation about the ethics of artificial intelligence beyond computer science and related fields to include other fields of scholarly endeavor, including the social sciences, humanities, and the professions (law, medicine, engineering, etc.), and
- it invites critical analysis of all aspects of—and participants in—the wide and continuously expanding artificial intelligence complex, from production to commercialization to consumption, from technical experts to venture capitalists to self-regulating professionals to government officials to the general public.

Ideally, handbooks combine stock-taking and genre-defining. Devoted to a field of inquiry as new and quickly-evolving as ethics of AI, this handbook falls closer to the forward-facing than to the literature-reviewing end of the spectrum. Mapping the existing discourse is important, also as the beginning of a crucial attempt to place current developments in historical context. At the same time, we recognized the need to leave room for flexibility as the contributors to this volume broke new ground, pursuing fresh approaches and taking on novel subjects. In the same spirit, this handbook operates with an inclusive and flexible conception of "artificial intelligence" that ranges from exploring normative constraints on specific applications of machine learning algorithms to reflecting on the (potential) status of AI as a form of consciousness with attendant rights and duties and, more generally still, to investigating the basic conceptual terms and frameworks necessary to understand tasks requiring intelligence, whether "human" or "AI."

Each chapter in this handbook aims to provide an original, critical, and accessible account of the current state of debate in its domain that will help to shape scholarly research and public discourse. We have welcomed forward-looking and ideas-driven contributions, to serve as catalysts for shaping the debate on the ethics of AI rather as mere stocktaking exercises. The chapters are

intended to function, individually and collectively, as lively, freestanding essays targeted at an international and interdisciplinary audience of scholars and interested laypersons. Each chapter also provides, at the end, a bibliography of about ten titles for readers who would like to read more deeply into the topic.

The handbook's inclusive and flexible approach to its subject matter is reflected in its roster of contributors, which includes authors from several countries and continents, ranging from emergent to established authorities and representing a wide variety of methodological approaches, areas of expertise, and research agendas. The handbook's content is similarly ambitious and diverse in scope and substance, covering a broad range of topics and perspectives. The handbook consists of five parts: I. Introduction & Overview, II. Frameworks & Modes, III. Concepts & Issues, IV. Perspectives & Approaches, and V. Cases & Applications.

Part I provides a general introduction to the subject (and field) of "artificial intelligence" within the context of research and discourse in related fields of technological innovation, laying an accessible yet nuanced foundation for the exploration of various normative frameworks for the critical analysis of AI. It also locates the "ethics" of artificial intelligence in relation to cognate fields of ethical inquiry (e.g., data ethics, information ethics, robot ethics, internet ethics), considering ways of conceptualizing it and its challenges (e.g., as a *sui generis* inquiry, as a form of applied ethics, or as traditional ethics in AI terms), distinguishing aspects within it (to the extent a taxonomy of this sort proves illuminating), and capturing some key substantive and formal features of the discourse.

Part II places the subject of this handbook, the ethics of AI, within the context of alternative frameworks for normative assessment and governance, including various institutional and procedural modes of implementation and dissemination. Questions raised in this Part include: "What distinguishes the ethics of AI from other normative frameworks and techniques, e.g., law, policy, regulation, governance?"; "How can ethics ground and inform legal constraints on (and regulatory guidance for) AI?"; "How does an ethics of AI navigate the possible tension between private commercial norms, on the one hand, and public norms, on the other?"; "How should ethical norms be generated and formulated, disseminated and implemented, and by whom?"; and "What is the role of the (self-)regulation of professional ethics, insofar as this enterprise is regarded as defining and enforcing a notion of good, sound, or 'professional' judgment?"

Part III tackles central concepts and issues that may serve as points of departure for reflecting on the ethical dimensions and challenges of artificial intelligence in general, cutting across technologies and applications, and in many cases across disciplines as well, ranging from the sources and types of bias in the production and application of AI research, to concerns about privacy in the collection and use of data, the potential effect of AI-driven "disruption" on labor markets and the future of work and on socio-economic life more broadly, the distinction between

"prediction" and "judgment," and the ethical status of AI-driven machines and its possible implications for human-machine interaction.

While a wide spectrum of disciplinary, national, and supranational perspectives is reflected throughout the handbook, Part IV homes in on a selection of methodological approaches and domestic or regional contexts. Early chapters in this Part capture the distinctive texture and salience of actual (or potential) discourse around ethics of AI in a range of disciplinary contexts, in an effort to illustrate—and to expand—the disciplinary scope of the scholarly and public debate about ethics of AI. The remaining chapters highlight the variety of discourses around ethics of AI in selected national and regional contexts, again to broaden and to diversify the dialogue about the normative dimensions of artificial intelligence as a global phenomenon, this time geographically and culturally.

Part V concludes the handbook by sharpening its focus to selected applications of artificial intelligence, without, however, treating them as *sui generis*, but instead in a way that fits into the handbook's overall ambition: to expand the conversation about the ethics of artificial intelligence from the specific to the general, from the superficial to the fundamental, and from the parochial to the contextual. Contributors here reflect on the ethical aspects of the design, dissemination, and use of AI-driven devices and tools today and in the future, along a broad spectrum of applications, in health care, law, immigration, education, transportation, the military, the workplace, smart cities, and beyond.

We are deeply grateful to the international and interdisciplinary group of scholars who signed on to this large-scale long-term project and somehow made the time to see it through to completion, among the flurry of activities and opportunities that mark the start of a new and momentous endeavor like the scholarly and public scrutiny of the ethics of artificial intelligence.

Markus D. Dubber, Frank Pasquale, and Sunit Das
August 2019