Journal of Artificial Intelligence Research 73 (2022) 619-631 Submitted 06/2021; published 02/2022

Viewpoint: Ethical By Designer - How to Grow Ethical Designers of Artificial Intelligence

Loïs Vanhée *(Corresponding author)* LOIS.VANHEE@UMU.SE
Umeå Universitet
Department of Computing Science
Linnaeus väg 49, 907 36 Umeå, Sweden

Melania Borit MELANIA.BORIT@UIT.NO
UiT The Arctic University of Norway
Norwegian College of Fisheries Science
Muninbakken 21, 9019 Tromsø, Norway

Abstract

Ethical concerns regarding Artificial Intelligence (AI) technology have fueled discussions around the ethics training received by AI designers. We claim that training designers for ethical behaviour, understood as habitual application of ethical principles in any situation, can make a significant difference in the practice of research, development, and application of AI systems. Building on interdisciplinary knowledge and practical experience from computer science, moral psychology and development, and pedagogy, we propose a functional way to provide this training.

1. Introduction

Against the backdrop of the challenges posed by ethical dilemmas of Artificial Intelligence (AI), spanning from bias in AI systems (Lee, 2018) to manipulation of human judgement (Henriksen, 2019), *virtue ethics*, understood as an approach to normative ethics that emphasizes moral character in contrast to approaches that emphasize duties and rules (deontology) or consequences of actions (consequentialism) (Carr, 2008; Hursthouse, 2017), becomes increasingly important in the debate around the impact AI will have on society. Virtue ethics gain attention as current tertiary education seems to fail in developing professional ethics and social responsibility skills (Chang, Shih, & Chang, 2020), codes of ethics are not drivers of ethical behaviour in moral exemplars in computing (Huff & Furchert, 2014), and developers' compliance with the principles set out in the various ethical guidelines is poor (McNamara, Smith, & Murphy-Hill, 2018). While moving away from preaching rules to focusing on cultivating the developers' character dispositions and moral attitude is a sensible advice (Harris, 2008), how to follow it is not straightforward, either for educators or for learners. We believe that an interdisciplinary approach integrating knowledge and experience from computer science, moral psychology and development, and pedagogy can provide a way for "broadening the scope of action, uncovering blind spots, promoting autonomy and freedom, and fostering self-responsibility" (Hagendorff, 2020) - that is, training for *ethical behaviour*, understood as habitual application of ethical principles in any situation (Treviño, Weaver, & Reynolds, 2006) (e.g., routinely record anonymisation procedures in data mining activities involving personal data).

Thus, to those interested in becoming the kind of AI systems developers that the society needs and to those willing to contribute with training such developers, we propose using the GEDAI framework - Growing Ethical Designers of Artificial Intelligence. This framework is based on an integrative curricular strategy that is consistent with approaches that make the ethical and value aspects explicit as a part of the design process (Shilton, 2013, 2014, 2018; Shilton & Anderson, 2017; Spiekermann, 2019). However, implementing our GEDAI framework in teaching and learning practices will mark a shift from achieving ethics for design(ers) (i.e., action-restriction through strict regulation of practice) (Dignum, 2019) to achieving *ethics by designers* - that is, empower AI systems developers to act self-responsibly in situations where morally relevant decisions have to be made. In the following, we describe this framework and the way in which we envisage its use.

2. Description of the GEDAI Framework

We have chosen to conceptualize our ideas of integrating teaching ethical behaviour within teaching AI practices in the form of a framework, that is, the GEDAI framework. The elements of the GEDAI framework and the relationship between them are visualized in Figure 1.

The GEDAI framework builds on four core *principles*:

A) Ethical Behaviour (EB) is a central concept in virtue ethics and it is rooted in the social condition and the human psyche (Rest, Bebeau, & Volker, 1986). As such, GEDAI proposes growing EB using advances in the domain of moral psychology and development. Within this domain, the Four Component Model of Moral Behavior (Narvaez & Rest, 1995) is the most studied and applied. This model introduces four dimensions along which individual moral ability and behavior can be grown: moral *sensitivity*, moral *judgement*, moral *motivation*, and implementation, also referred to as moral *action*. Moral sensitivity captures the ability to observe the morally-sensitive aspects from raw observations (e.g., emotions, ramifications of decisions). Moral judgement captures the ability to decide upon the most sensitive course of conduct when facing a morally loaded situation. Moral motivation captures the inclination of fulfilling moral duties, over other dimensions. Moral action captures the ability for actually implementing a moral decision, including performative ability and psychological resilience.

B) Teaching ethical behaviour to AI systems developers can be operationalised using Intended Learning Outcomes (ILOs). Being statements about what a learner will achieve upon successful completion of an instructional unit (IU), ILOs are expressed from the learners' perspective and are measurable, achievable, and assessable. As such, GEDAI advocates for defining specific ILOs for learning ethical behaviour.

C) Ethical behaviour is a transferable skill and, as such, it can be integrated with the practices of teaching AI. Ethics education in a professional context has tended to follow one of three basic models: a dedicated ethics course offered by a humanities & social science department as a requirement or an elective; a dedicated ethics course within the discipline as a requirement or an elective; or ethics content integrated into one or more core courses. We believe that teaching ethical behaviour can and should be smoothly integrated in regular AI IUs, be they individual sessions, modules, courses, or programs, as integration as a strategy for developing transferable skills is proven to be more effective in higher education

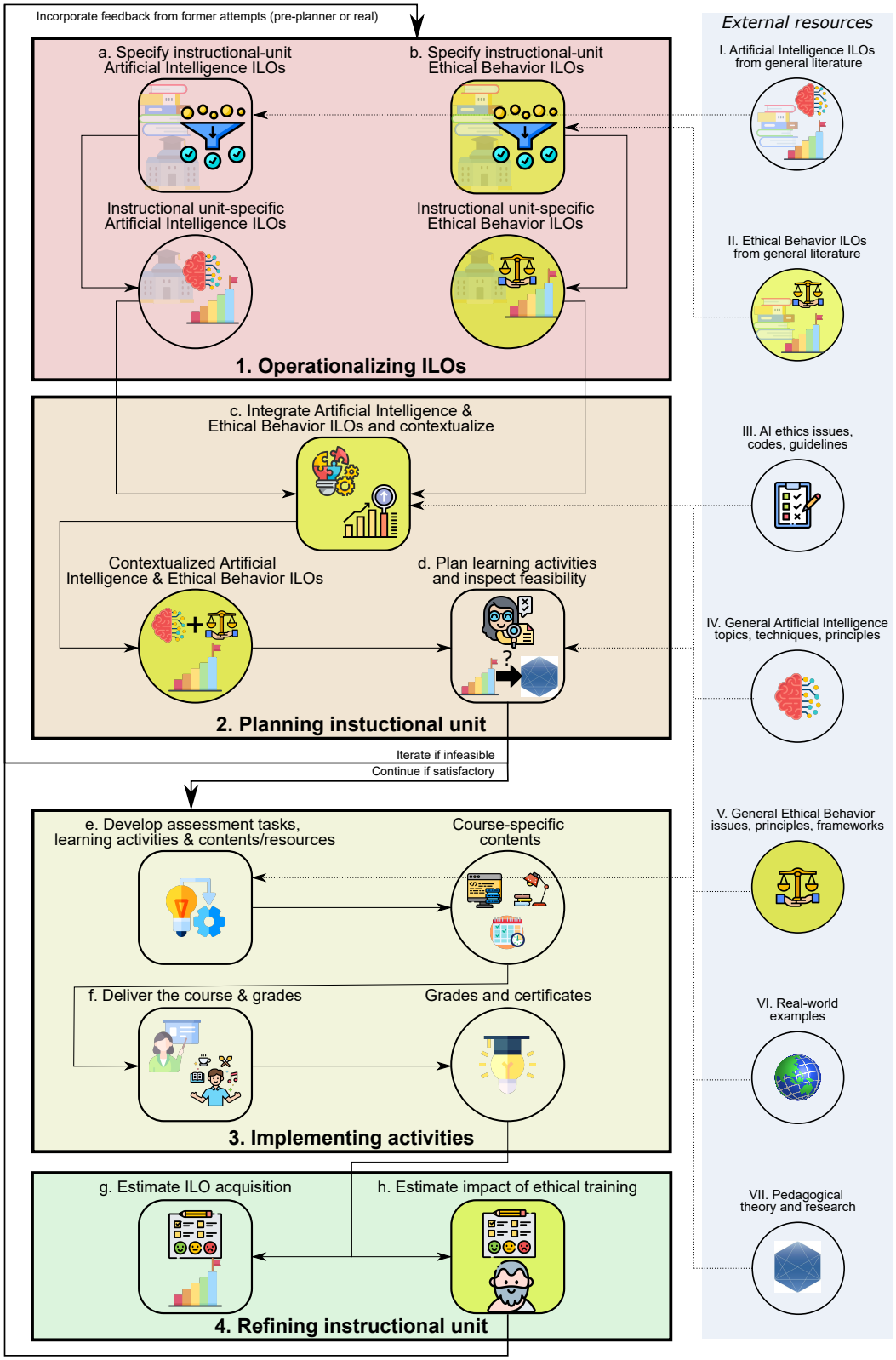

Figure 1: GEDAI - Growing Ethical Designers of Artificial Intelligence framework. Squares represent actions; circles represent inputs/outputs. The yellow highlighted items represent the distinctive elements that integrate teaching ethical behaviour in the AI teaching and learning process. Intra- and interphases loops are not displayed.

as it is more representative of the real-life application of skills in the workplace (Cottrell, 2001). However, the GEDAI framework is still applicable in cases where a dedicated ethics course within the discipline is used.

D) Learners construct ethical behaviour, meaning, habits, and expertise through relevant learning activities, while teachers' task is to set up a learning environment that supports these learning activities. As such, GEDAI uses the *constructive alignment* educational principle (Biggs & Tang, 2011) to the design of IUs that integrate AI and ethical behaviour teaching. According to this approach, ILOs are identified, then assessment tasks are designed to measure the attainment of these ILOs, followed by planning learning activities and choosing content required to support these learning activities.

The main elements of the teaching and learning process depicted by the GEDAI framework are phases, actions, and inputs/outputs. These *phases* are: 1) Operationalizing ILOs; 2) Planning of IU; 3) Implementing activities; 4) Refining IU. Whereas being presented here in a sequential order for facilitating understanding, these actions and phases are to be *undertaken in loops*, where actions/phases can overlap or a specific action/phase can trigger a revision of the previous one(s) (i.e., intra- and interphases loops). For example, once a concrete activity plan is defined at the end of Phase 2 action d, framework users are invited to reconsider the set of ILOs to be covered and that were defined in Phase 1 actions a and b.

In **Phase 1 (Operationalizing ILOs)**, the framework user has the task to specify ILOs suitable for the respective IU starting from higher level ILOs (actions a and b in Figure 1). In the case of AI, such high level outcomes can be found in program descriptions and educational plans issued by the Association for Computing Machinery and the IEEE Computer Society, or similar documents. In the case of EB, besides program descriptions, educational plans, and similar documents, scientific publications bring a significant contribution; (see, for example, Martin, Huff, Gotterbarn, & Miller, 1996; Martin & Weltz, 1999; Martin & Huff, 1997; Werth, 1997; Harris Jr, Davis, Pritchard, & Rabins, 1996; Callahan, 1980; Huff & Frey, 2005; Frey, 2010; Lennerfors, Laaksoharju, Davis, Birch, & Fors, 2020).

The task to specify AI ILOs (action a) using Bloom's taxonomy is already a common practice (see, for example, Manaris, Wainer, Kirkpatrick, Stalvey, Shannon, Leventhal, Barnes, Wright, Schafer, & Sanders, 2007) or Sobral (2021). However, in the case of EB ILOs (action b), this specification of IU-specific EB ILOs from the general literature is not clear. Thus, here we propose using the *Integrative Ethical Education Model* (Narvaez & Lapsley, 2008; Narvaez & Bock, 2014), further operationalised and supplemented with examples of assessment and activities in the *Ethical Expertise Model* (Narvaez, 2009; Narvaez & Lies, 2009; Narvaez & Endicott, 2009; Narvaez & Bock, 2009). These educational models build on the Four Component Model of Moral Behaviour, explained above in Principle A, and are based on evidence that such behaviour can be fostered by training *ethical expertise* (Huff, 2014; Narvaez, 2010), which is best gained through a novice-to-expert approach that moves through several stages of instruction while blending well-educated intuitions and good reasoning.

In **Phase 2 (Planning instructional unit)**, the framework user has the difficult task to integrate the two sets of operationalised ILOs and specify ILOs for the respective IU, and contextualize these ILOs (action c). GEDAI chooses to use an *integrative strategy* to teaching ethical behaviour as a transferable skill (i.e., skills are developed and taught explicitly within

the core discipline with equal emphasis given to transferable skills and technical abilities), as opposed to embedding (i.e., no direct reference is made to developing transferable skills and the emphasis is on promoting the development of technical 'know-how') or bolting-in (i.e., skills are developed independently of the core discipline, enabling the explicit development of learners' transferable skills) (Chadha, 2006). This integrative approach is considered to be more effective in higher education as it is more representative of the real-life application of skills in the workplace (Cottrell, 2001).

Several inputs play a role in action c (items III-VII in Figure 1), and the user has to be skilled in combining knowledge from various domains: AI in theory and in practice - for example, contents of well established handbooks, such as Russell and Norvig (2002), as well as new trends, such as autonomous cars; pedagogy - for example, how to establish the position of the IU at mastery, familiarity, or exposure level (McDonald & McDonald, 1999), or the use of the Didactic Planning Diamond (Bjørndal & Lieberg, 1978)); ethics in general - for example, classic ethical theories, such as, deontic, utilitarian, and virtue ethics; AI ethics - for example, the guidelines drawn by the High-level Expert Group on Artificial Intelligence at the European Commission (HLEG, 2019); and the presence of AI in the real-world - for example, a robot-dog used by the police being discarded after deployment, due to its looks causing major public disapproval (Zaveri, 2021).

At the end of Phase 2, the framework user has to make a plan of learning activities that has to undergo a feasibility check (action d) during which it is assessed whether the planned activities are aligned with the ILOs and are compatible with administrative constraints (e.g., teacher time, learner time, rooms availability) and with other teaching activities, in order to avoid *non-productive* repetition (e.g., accidentally teaching exactly the same content about algorithmic bias in exactly the same way in several IUs) and to cover blind spots (e.g., two teachers assuming that the other teacher covers data anonymization issues). While the current version of the GEDAI framework makes this planning action explicit, further specifications of the framework can go into the details of what learning activities and teaching methods to use in order to achieve the integrated AI and EB ILOs or of what assessment tasks are the most suitable for these specific ILOs.

In **Phase 3 (Implementing activities)**, the framework user carries on teaching as in the case of any other IU, with the mention that input V becomes relevant when the IU is an AI course in which ethics content is being integrated, in contrast with input IV that becomes relevant when the IU is a dedicated ethics course within the AI discipline. Considering that the output of action c is a set of innovative integrated AI and EB ILOs, the framework user has to be aware that innovative assessment tasks have to be formulated in action f, to assess the achievement of these ILOs. For example, a panel might be formed to construct a skill measurement approach that was targeted to AI and based on the ethical behaviour literature and practice and customized for the specific ILOs[1].

In **Phase 4 (Refining instructional unit)**, the framework user performs the usual action of estimating ILOs acquisition by the learners (action g) through, for example, correlating grades with learners' feedback. Regardless of whether ethical skills acquisition was graded or not in Phase 3, in Phase 4, in addition to action g, we propose performing an estimation of the impact of ethical training (action h), which usually also involves collecting

1. We thank one of the anonymous reviewers for this suggestion.

baseline data (Rhode Island Department of Education & the National Center for the Improvement of Educational Assessment, 2013). This estimation intends to gauge the ethical growth of the learners and, as such, the efficiency of the educational activity in terms of growing ethical development.

If the framework user has chosen to use the Ethical Expertise Model (Narvaez & Bock, 2014) in action b, to our knowledge, there is no assessment tool that holistically addresses the four components of moral behaviour as described in Phase 1 above, except a self-scoring instrument developed for dental education (Chambers, 2011) and a questionnaire designed for veterinary students (Verrinder & Phillips, 2014b). However, tools exist for assessing all of the four individual components. These tools can be general or profession-specific. As examples of general ones, we can mention the Defining Issues Test, which is argued to be usable for assessing moral judgement abilities at group level (Rest & Narvaez, 1994) - for an overview of such instruments for assessing moral thinking see Goldsmith, Burton, Dueber, Goldstein, Sampson, and Toland (2020), or the Professional Role Orientation Inventory (Bebeau, Born, & Ozar, 1993), which has been used to measure moral motivation (Bebeau & Monson, 2014). As examples of profession-specific tools, we can mention the Dental Sensitivity Test, which includes stimulus materials and scoring procedures to measure an individual's ability to recognize the ethical issues often hidden within the dentist's professional problems (Bebeau, Rest, & Yamoor, 1985), or the Veterinary Defining Issues Test, designed for veterinary education (Verrinder & Phillips, 2014a). For examples of several other profession-specific tools see Bebeau and Monson (2014), You, Maeda, and Bebeau (2011), You and Bebeau (2012). The user of the GEDAI framework can explore how such profession-specific tools can be adapted for AI education.

The *strength* of the GEDAI framework lies in the following. GEDAI focuses on teaching and learning ethical *behaviour*, which can be more straightforwardly embedded in daily life than other ethics-related skills. This approach seeks to grow *practical techno-ethical competent* learners, that is, technically-able persons with the habit of using ethical skills when producing concrete technical contents (e.g., growing the habit of relating "training data" to "assessing bias" rather than jumping to counting the layers of the neural network). The GEDAI learners will be able of integrating ethics beyond just writing the optional ethics appendix after the implementation of a system and beyond fault-finding criticism with no pragmatic evaluation of solutions and alternatives. The GEDAI framework uses an integrative strategy to teaching ethics (i.e., teaching ethical skills explicitly within the core discipline and placing the same amount of emphasis on the development of these skills as on developing technical abilities), an approach which is considered to be more effective as it is more representative of the real-life application of skills in the workplace (Cottrell, 2001). At the same time, GEDAI makes explicit all the necessary steps to achieve this integration, thus being more concrete or operational than other available solutions. The GEDAI framework provides high adaptability by relying on meso-level "containers" that can be filled in with specific content depending on the context (e.g., what are the hot topics of the moment). This feature ensures the longer-term effectiveness of the framework, as it can adapt to the moving landscape of ethical issues and deontological rules. That being said, possibly the strongest aspect of the GEDAI framework lies in its potential to facilitate the training of learners that have both technical know-how and the necessary ethical skills

to *use* this know-how in the "right" way; for example, when engaging with the task of defining the system's objective function, the designer taught under the GEDAI framework will have the habit of remaining vigilant about direct, indirect, and unexpected involved stakeholders; the automatic routine of verifying that the system's behavior derived from this function complies with all stakeholders; and the technical mastery of the mathematical arcanes for this compliance to be achieved by the system. However, as in the case of any other pedagogical tool or model, GEDAI only provides a foundation for organizing teaching and learning environments that maximize the chances for practical ethical expertise to be acquired, but cannot guarantee that this expertise will be demonstrated on the field.

We believe that giving our students the best chance to learn is of paramount importance when performing our activities as teachers.

As a final note, we have to mention that based on our own teaching experience and on numerous discussions with colleagues and learners, we are aware of the challenges posed by including yet another layer in the complex fabric of what has to be taught to a specific set of learners. Multi-cultural awareness, 21st Century skills (i.e., collaboration, communication, creativity, and critical thinking), entrepreneurship, or ethics, all are considered essential for today's learners. Contrary to the general feeling of some educators, how to make space for all these in our teaching without having to remove extensive AI disciplinary learning is a skill that we have to grow ourselves as teachers. However, this is not possible without the collaboration of the education leadership, the administration, the other teaching staff, and of the learners themselves. The leadership has to prioritize teachers' growth time as educators over minimizing teaching costs, as quality student-oriented teaching adapted to the needs of our times requires more time and resources than the usual frontal teaching. The administration has to be able to adapt to the needs of implementing this teaching (e.g., adapted physical rooms, customized Learning Management Systems, educational offers that adapt to the issues relevant in the society at a given time). The other teaching staff has to be open to have the same approach to their teaching as they have to their research. Thus, making use of learning analytics (Gašević, Kovanović, & Joksimović, 2017) and latest research in pedagogy should be the norm, not the exception. The learners have to see themselves as co-creators of value for themselves and for society from the moment they enter an educational program and not as dormant entities that will be activated after finishing a degree.

3. Who Can Use GEDAI

Learning professional ethics is acknowledged to benefit learners and professionals (Bebeau & Monson, 2008). Thus, we envisage several user groups for the GEDAI framework: instructional units designers, industry, individual learners, researchers, and grant funders.

Designers of instructional units (teachers, course coordinators, program directors) who want to create a unit from scratch or update an existing unit can use the framework as a complementary tool to any other tools that are out there for instructional design (see, for example, Branch & Kopcha, 2014). Moreover, they can use the framework to structure their critical reflection on choices to be made when integrating ethics in IU design and implementation. Action d will provide an overview of what skills are taught across units and avoid unproductive repetition or gaps. Growing ethical behaviour can be seen as a

transferable skill that can be taught across several IUs, like, for example, collaboration or cross-cultural communication. Since GEDAI is explicit about what external resources to feed into learning activities (items I-VII), it contributes to aligning activities in AI education with developments that happen in the real-world, thus aligning these activities with the needs of society and of learners. In terms of achieving the desired ethical behaviour, the framework helps with suggesting operational ILOs for this domain. Consistent application of the framework will help learners move on the continuum of moral behaviour expertise. As further aid, designers of IUs can build up on experiences of practitioners from other fields who have an integrative approach to training ethical behaviour (see, for example, Huff & Furchert, 2014; Huff, 2014).

Industry stakeholders interested in improving the capabilities of their workforce and in fulfilling their corporate social responsibility can use the framework in a similar way as the IU designers. The GEDAI framework provides a structure that can be customized for the individual needs of the company (e.g., the needs of data analytics vs. robotics company). At the same time, using the framework may reduce the transition costs towards ethical practices by allowing a smoother, incremental uptake into practice of the AI ethics guidelines by their employees. An interesting effect of using the framework in such an industry setting is that, since ethical behaviour is a transferable skill, then other areas of the company activity might benefit from increased ethical practice due to the habits developed by the employees. However, this transferability has to be considered with caution, as it is highly dependant on the quality of the learning underlying the respective skill, on the context in which the skill was learned, and on the context in which it has to be transferred (Sasson & Dori, 2012). Moreover, according to Huff (2014), a distinction has to be made between general skills that support moral behaviour (e.g., self-regulation, self-control, emotion regulation) and specialized skills that are applicable to particular domains and technical activities (e.g., listing common sources of visible discriminations made by neural network applications; identifying sources of hazard and develop appropriate fail safe procedures for a heavy robot operated in a confined environment), as the latter ones might be more difficult to employ in a different domain or activity.

Learners, both those learning by themselves and those enrolled in formal education, can use the GEDAI framework as an awareness-raising tool. By examining the elements of the framework, learners become aware of what is necessary to include in their self-growth process (e.g., various external resources), of the opportunity of learning ethics as a transferable skill together with learning the technical skills (much like learning statistics, for example), and of the need to estimate own ethical growth (action h) and measure their progression in becoming a responsible professional.

Funders of projects in education (e.g., the European Union) can use the framework for developing concrete funding programs and/or calls that promote the blending of ethics within operational skills, rather than as yet another requirement that ends up being presented as an appendix, while retaining the flexibility to incorporate what is relevant in the society at a certain point. The implementation of projects funded through such programs/calls would contribute to the growth of ethical behaviour skills in general and within the AI domain in a set-up that is more relevant to real-life and work-place situations.

Researchers interested in curriculum assessment can use the GEDAI framework for specifying content analysis codes suited for exploring how ethics are integrated in the cur-

rent AI teaching practices or from a historical point of view. Those interested in the theoretical and methodological development of the field of teaching AI ethics can build upon the framework and possibly conceptualize/formalize it further. Researchers can use GEDAI for further framing the various contributions that can be made in teaching AI ethics (e.g., pairing activities and grading methods to ILOs).

4. Conclusions

Integrating abstract ethical recommendations and technical implementations is not a trivial task. Embedding ethical behaviour at the core of teaching and learning AI courses can help, and we propose drawing on expertise in computer science, moral psychology and development, and pedagogy to crack this hard nut. Being able to develop AI focused on social good now as well as in the future requires growing developers who behave ethically as a habit, even in the absence of an explicit set of rules, duties, or imperatives. Climbers are safe when venturing up a rock by performing safety habits that they learned early in their training: double back their own harness buckle, check the belay system, or go through the starting commands. Likewise, GEDAI-grown developers will be ethical when designing an AI system by performing ethics habits they learned early in their training: consider the misuse of the system by unintended stakeholders, such as hackers; assess the impact of the optimisation algorithm on the well-being of it users; compare how alternative implementations of a neural network propagate discriminatory bias; or, despite systemic pressure, voice concerns of overlooking critical ethical concerns for cheaper software. Such a habit development in the GEDAI designers will have a long-ranging positive influence on the impact AI can have on society.

Acknowledgements

The first author acknowledges the support of the Knut and Alice Wallenberg Foundation (project number 570080103), of the project AutograIde "A Student-Driven Interdisciplinary Hackathon on Whether and How to Automate Grading & Assessment" (project number 570002260, The Rådet för Artificiell Intelligens, Umeå University, Sweden), and of the project GEDAI "Growing Ethical Designers of Artificial Intelligence", Umeå University, Sweden. Special thanks go to Bas Testerink, Frank Dignum, and Maarten Jensen for their comments on the layout of Figure 1 and for `https://www.flaticon.com/` for some of the icons used in the image. The second author acknowledges the support of the project AFO-JIGG "Service design thinking to improve welfare and product quality in the Norwegian small-scale fishing fleet" (project number 302635, Research Council of Norway) and of the CRAFT Lab - Knowledge Integration and Blue Futures, UiT The Arctic University of Norway.

References

Bebeau, M. J., Born, D. O., & Ozar, D. T. (1993). The development of a professional role orientation inventory.. *The Journal of the American College of Dentists*, *60*(2), 27–33.

Bebeau, M. J., & Monson, V. E. (2008). *Guided by theory, grounded in evidence: A way forward for professional ethics education.*

Bebeau, M. J., & Monson, V. E. (2014). A theoretical and evidence-based approach for designing professional ethics education. In *Handbook of moral and character education*, pp. 507–534. Routledge.

Bebeau, M. J., Rest, J. R., & Yamoor, C. M. (1985). Measuring dental students' ethical sensitivity. *Journal of Dental Education*, *49*(4), 225–235.

Biggs, J. B., & Tang, C. (2011). *Teaching for quality learning at university.* McGraw Hill education (UK).

Bjørndal, B., & Lieberg, S. (1978). *Nye veier i didaktikken?: en innføring i didaktiske emner og begreper.* Aschehoug.

Branch, R. M., & Kopcha, T. J. (2014). Instructional design models. In *Handbook of research on educational communications and technology*, pp. 77–87. Springer.

Callahan, D. (1980). Goals in the teaching of ethics. In *Ethics teaching in higher education*, pp. 61–80. Springer.

Carr, D. (2008). *Character education as the cultivation of virtue.* Routledge New York, NY.

Chadha, D. (2006). A curriculum model for transferable skills development. *Engineering Education*, *1*(1), 19–24.

Chambers, D. W. (2011). Developing a self-scoring comprehensive instrument to measure rest's four-component model of moral behavior: The moral skills inventory. *Journal of dental education*, *75*(1), 23–35.

Chang, J.-C., Shih, H.-F., & Chang, K.-L. (2020). A Research on the Training Status of EECS Students' Core Competency in University of Science and Technology. In *International Conference on Industrial Engineering and Engineering Management*, pp. 1069–1072. IEEE.

Cottrell, S. (2001). *Teaching study skills and supporting learning.* Palgrave Basingstoke.

Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way.* Springer Nature.

Frey, W. J. (2010). Teaching virtue: Pedagogical implications of moral psychology. *Science and engineering ethics*, *16*(3), 611–628.

Gašević, D., Kovanović, V., & Joksimović, S. (2017). Piecing the learning analytics puzzle: A consolidated model of a field of research and practice. *Learning: Research and Practice*, *3*(1), 63–78.

Goldsmith, J., Burton, E., Dueber, D. M., Goldstein, B., Sampson, S., & Toland, M. D. (2020). Assessing Ethical Thinking about AI. In *Proceedings of the AAAI Conference on Artificial Intelligence, 34(09)*, pp. 13525–13528.

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines*, *30*(1), 99–120.

Harris, C. E. (2008). The good engineer: Giving virtue its due in engineering ethics. *Science and Engineering Ethics*, *14*(2), 153–164.

Harris Jr, C. E., Davis, M., Pritchard, M. S., & Rabins, M. J. (1996). Engineering ethics: what? why? how? and when?. *Journal of Engineering Education*, *85*(2), 93–96.

Henriksen, E. E. (2019). Big data, microtargeting, and governmentality in cyber-times. The case of the Facebook-Cambridge Analytica data scandal. Master's thesis.

HLEG (2019). Ethics guidelines for trustworthy AI, High-Level Expert Group on AI. Tech. rep..

Huff, C. (2014). From meaning well to doing well: Ethical expertise in the GIS domain. *Journal of Geography in Higher Education*, *38*(4), 455–470.

Huff, C., & Frey, W. (2005). Moral pedagogy and practical ethics. *Science and Engineering Ethics*, *11*(3), 389–408.

Huff, C., & Furchert, A. (2014). Toward a pedagogy of ethical practice. *Communications of the ACM*, *57*(7), 25–27.

Hursthouse, R. (2017). Virtue Ethics in Stanford Encyclopedia of Philosophy, Metaphysics Research Lab, Palo Alto, CA..

Lee, D. (2018). Amazon scrapped 'sexist AI'tool, BBC News, October 10..

Lennerfors, T. T., Laaksoharju, M., Davis, M., Birch, P., & Fors, P. (2020). A Pragmatic Approach for Teaching Ethics to Engineers and Computer Scientists. In *Frontiers in Education Conference*, pp. 1–9. IEEE.

Manaris, B., Wainer, M., Kirkpatrick, A. E., Stalvey, R. H., Shannon, C., Leventhal, L., Barnes, J., Wright, J., Schafer, J. B., & Sanders, D. (2007). Implementations of the CC' 01 human-computer interaction guidelines using Bloom's taxonomy. *Computer Science Education*, *17*(1), 21–57.

Martin, C. D., & Huff, C. (1997). A conceptual and pedagogical framework for teaching ethics and social impact in computer science. In *Proceedings Frontiers in Education 1997 27th Annual Conference. Teaching and Learning in an Era of Change*, Vol. 1, pp. 479–483. IEEE.

Martin, C. D., Huff, C., Gotterbarn, D., & Miller, K. (1996). A framework for implementing and teaching the social and ethical impact of computing. *Education and Information Technologies*, *1*(2), 101–122.

Martin, C. D., & Weltz, E. Y. (1999). From awareness to action: Integrating ethics and social responsibility into the computer science curriculum. *ACM SIGCAS Computers and Society*, *29*(2), 6–14.

McDonald, M., & McDonald, G. (1999). Computer science curriculum assessment. *ACM SIGCSE Bulletin*, *31*(1), 194–197.

McNamara, A., Smith, J., & Murphy-Hill, E. (2018). Does ACM's code of ethics change ethical decision making in software development?. In *Proceedings of the 2018 26th ACM joint meeting on european software engineering conference and symposium on the foundations of software engineering*, pp. 729–733.

Narvaez, D. (2009). *EthEx: Nurturing character in the classroom. Book Four: Ethical Action.* Chapel Hill, NC: Character Development Publishing.

Narvaez, D. (2010). Moral complexity: The fatal attraction of truthiness and the importance of mature moral functioning. *Perspectives on Psychological Science*, *5*(2), 163–181.

Narvaez, D., & Bock, T. (2009). *EthEx: Nurturing character in the classroom. Book Two: Ethical Judgment*. Chapel Hill, NC: Character Development Publishing.

Narvaez, D., & Bock, T. (2014). *Developing ethical expertise and moral personalities*. Routledge New York, NY.

Narvaez, D., & Endicott, L. (2009). *EthEx: Nurturing character in the classroom. Book One: Ethical Sensitivity*. Chapel Hill, NC: Character Development Publishing.

Narvaez, D., & Lapsley, D. (2008). Teaching moral character: Two alternatives for teacher education. *The Teacher Educator*, *43*(2), 156–172.

Narvaez, D., & Lies, J. (2009). *EthEx: Nurturing character in the classroom. Book Three: Ethical Motivation*. Chapel Hill, NC: Character Development Publishing.

Narvaez, D., & Rest, J. (1995). The four components of acting morally. *Moral behavior and moral development: An introduction*, *1*(1), 385–400.

Rest, J. R., Bebeau, M., & Volker, J. (1986). *An overview of the psychology of morality*. Praeger New York, NY.

Rest, J. R., & Narvaez, D. (Eds.). (1994). *Moral development in the professions: Psychology and applied ethics*. Psychology Press.

Rhode Island Department of Education & the National Center for the Improvement of Educational Assessment, I. (2013). Using baseline data and information to set slo targets. https://www.ride.ri.gov/Portals/0/Uploads/Documents/Teachers-and-Administrators-Excellent-Educators/Educator-Evaluation/Online-Modules/Using-Baseline-Data-and-Information-Guidance.pdf.

Russell, S., & Norvig, P. (2002). *Artificial intelligence: a modern approach* (2nd edition edition). Prentice Hall.

Sasson, I., & Dori, Y. J. (2012). Transfer skills and their case-based assessment. In *Second international handbook of science education*, pp. 691–709. Springer.

Shilton, K. (2013). Values levers: Building ethics into design. *Science, Technology, & Human Values*, *38*(3), 374–397.

Shilton, K. (2014). This is an intervention: Foregrounding and operationalizing ethics during technology design. In *Emerging pervasive information and communication technologies (PICT)*, pp. 177–192. Springer.

Shilton, K. (2018). Values and ethics in human-computer interaction. *Foundations and Trends® in Human–Computer Interaction*, *12*(2).

Shilton, K., & Anderson, S. (2017). Blended, not bossy: Ethics roles, responsibilities and expertise in design. *Interacting with computers*, *29*(1), 71–79.

Sobral, S. R. (2021). Bloom's Taxonomy to improve teaching-learning in introduction to programming. *International Journal of Information and Education Technology*, *11*(3).

Spiekermann, S. (2019). *Ethical IT innovation: A value-based system design approach*. Auerbach Publications.

Treviño, L. K., Weaver, G. R., & Reynolds, S. J. (2006). Behavioral ethics in organizations: A review. *Journal of management*, *32*(6), 951–990.

Verrinder, J. M., & Phillips, C. J. (2014a). Development of a moral judgment measure for veterinary education. *Journal of veterinary medical education*, *41*(3), 258–264.

Verrinder, J. M., & Phillips, C. J. (2014b). Identifying veterinary students' capacity for moral behavior concerning animal ethics issues. *Journal of veterinary medical education*, *41*(4), 358–370.

Werth, L. (1997). Getting started with computer ethics. In *Special Interest Group on Computer Science Education*, Vol. 29, pp. 1–5.

You, D., & Bebeau, M. J. (2012). Gender difference in ethical abilities of dental students. *Journal of dental education*, *76*(9), 1137–1149.

You, D., Maeda, Y., & Bebeau, M. J. (2011). Gender differences in moral sensitivity: a meta-analysis. *Ethics & Behavior*, *21*(4), 263–282.

Zaveri, M. (2021). N.Y.P.D. Robot Dog's Run Is Cut Short After Fierce Backlash. The New York Times, April 28, 2021. https://www.nytimes.com/2021/04/28/nyregion/nypd-robot-dog-backlash.html.