Steps to Designing AI-Empowered Nanotechnology:

A Value Sensitive Design Approach

*Steven Umbrello**

Advanced nanotechnology promises to be one of the fundamental transformational emerging technologies alongside others such as artificial intelligence, biotechnology, and other informational and cognitive technologies. Although scholarship on nanotechnology, particularly advanced nanotechnology such as molecular manufacturing has nearly ceased in the last decade, normal nanotechnology that is building the foundations for more advanced versions has permeated many industries and commercial products and has become a billion dollar industry. This paper acknowledges the socialtechnicity of advanced nanotechnology and proposes how its convergence with other enabling technologies like AI can be anticipated and designed with human values in mind. Preliminary guidelines inspired by the Value Sensitive Design approach to technology design are proposed for molecular manufacturing in the age of artificial intelligence.

I. Reinvigorating Nanoethics

Atomically precise manufacturing (APM) is one form of advanced nanotechnology that falls within the larger category of molecular manufacturing. This theoretical mode of manufacturing was developed in greater depth in K. Eric Drexler's 1986 APM book *Engines of Creation*. However, the concept of APM in theory dates to Richard Feynman's 1959 talk 'There's Plenty of Room at the Bottom.' Today, the majority of nanotechnology R&D is not APM per se, but instead is technology involving simpler nanometer-scale processes; this is sometimes referred to as 'normal nanotechnology'.[1] Many nanotechnology researchers likewise doubt the feasibility of APM and instead favor research on more directly promising nanotechnology directions. Despite these doubts, current investments and national interests towards the development of APM warrant investigations into how we can ensure the concept, and its convergences with other technologies, is as beneficial to humanity as possible by intervening at the design stages and incorporating the relevant values necessary to achieve a desired end.

There is a difficulty, however, in evaluating advanced nanotechnology per se, that is that nanotechnology is part of a converging set of transformative technologies such as biotechnology, information technologies and cognitive technologies like artificial intelligence. This muddies the waters in prescribing a rigid set of values, principles or positive governance structures to its development because it is hard to demarcate discrete boundaries between nanotechnologies and these others, despite its potentially transformative impacts. This co-constructive convergence of technologies can still be guided towards beneficial ends. The difficulty of guiding technological design and process should not be confused with technological determinism, stakeholders can and should engage with the design and development processes of technologies that involve them.

It, of course, may not be feasible to account for or engage with all the processes, variable or values implicated in the design of APM, however, certain conceptual steps can be taken and adopted by design

* Steven Umbrello, Managing Director at the Institute for Ethics and Emerging Technologies and graduate student at the University of Turin (Consorzio FINO), Italy. For correspondence: <steve@ieet.org>. This paper is built upon the author's previous work: Steven Umbrello 'Atomically Precise Manufacturing and Responsible Innovation: A Value Sensitive Design Approach to Explorative Nanophilosophy' (2019) 10 Int J Technoethics 2, 1-21

1 Donal O'Mathuna, *Nanoethics: Big Ethical Issues with Small Technology* (Continuum 2009)

teams to nudge their design flows towards beneficial outcomes. To this end, this short *ethics in practice* piece offers a modular and reflexive set of guidelines that can provide policy experts, design teams, industry leaders and even ethicists a way of conceptualising the way advanced nanotechnology design can be confronted that accounts for the values of those it may impact and how those values can be put into practice.

II. A Value Sensitive Design (VSD) Approach

In the interest of developing these guiding principles, or perhaps better termed as *framing tools*, I avoid discussion of what constitutes APM, the debate over its feasibility, as well as expert projections for when we can expect to see the first instantiation of APM technologies. Simply put, the literature that discuss APM and how it functions envisions impacts on an astronomical scale.[2] It has even been suggested that the safe development of APM, given that misaligned development could result in existential catastrophe, is contingent on the use of artificial intelligence (even artificial superintelligence). Regardless, the current nanotechnology research is laying the path for eventual APM systems to emerge. However, first a brief overview of what VSD is and how it functions is warranted so that the tools used to frame APM do not seem ad hoc, but rather integrative.

The VSD methodology emerged in the field of human-computer interaction and in particular from the realisations that two principal values were missing from the design process in technological innovation, user autonomy and freedom from bias.[3] Hence, VSD's intention is to ensure that designers include the values held by stakeholders in the early design phases to not only produce a satisfactory product for the stakeholder, but one that ensures that human values are advanced in technological innovation.[4]

The priority on the subsumption of human values is a critical given that the domain of converging technologies, as harmony between individual, corporate, and societal values is often lacking in facour of financial and/or socio-political gain.[5] Hence, this warrants a design space that considers the values of all stakeholders that the technology may potentially impact. VSD makes explicit claims to help foster a participatory space in which stakeholders can communicate both their values and desires.[6]

The VSD framework is predicated on the presumption that technology is something that is value-laden and thus is of significant ethical importance. The approach puts considerable emphasis on the human values of freedom, autonomy, privacy, and equality given that they were the values distilled both during conceptual investigations as well as empirical stakeholder elicitations.[7] Each of these values has the potential to be limited by technology and thus must be taken into account during the design of technologies. It focuses on the values of stakeholders and how those values can be reconciled with design and engineering limitations and constraints. Instead of the conventional way of appraising a technologies moral status, ie, how it is placed, used, and construed in a societal context, VSD aims determine the impact that technology has on the moral landscape. It aims to determine the values of stakeholders and integrate those values early on and throughout the design process. Similarly, the approach does not seek to revolutionise the engineering practices of designers in such a way that requires unique or burdensome requirements; instead, VSD is an instrument that proposes ways of changing existing design and engineering methods in such a way as to include stakeholder values and to adapt itself to the already engaged in engineering environments and practices. This is of particular importance to policy makers and industrial leaders in terms of increasing the approach's social acceptance.

---

2 Steven Umbrello and Seth Baum, 'Evaluating Future Nanotechnology: The Net Societal Impacts of Atomically Precise Manufacturing' (2018) Futures 100, 63–73

3 Alan Borning and Michael Muller, 'Next Steps for Value Sensitive Design' (SIGCHI Conference on Human Factors in Computing Systems, Austin, 5 - 10 May, 2012) <https://dl.acm.org/citation.cfm?doid=2207676.2208560> accessed 15 July 2019

4 Batya Friedman et al, 'Value Sensitive Design and Information Systems' in Neelke Doorn et al (eds), *Early Engagement and New Technologies: Opening up the Laboratory* (Springer 2013) 55 - 95

5 Langdon Winner, 'Do Artifacts Have Politics?' (2003) Technol. Futur. 2003, 109 (1), 148–164. https://doi.org/10.2307/20024652

6 Steven Umbrello, 'Atomically Precise Manufacturing and Responsible Innovation: A Value Sensitive Design Approach to Explorative Nanophilosophy' (2019) International Journal of Technoethics 10

7 Batya Friedman and Peter Kahn, 'Human Values, Ethics, and Design' in Julie Jacko (eds) *The Human-Computer Interaction Handbook* (CRC Press 2003) 1177–1201

III. Framing Considerations for Safe Nano-Futures

Framing is a way of envisioning ways that technological developments and potential socio-ethical and cultural issues can rise given a particular technology's development. Framing certain technological design flows (open design avenues) provides designers with principled ways to make informed design decisions.[8] The social sciences have an illustrious background with eliciting stakeholders in different contexts as well as determining their values, interests, and preferences[9] such as the use of *Envisioning Cards.*[10] The following elements are preliminary framing considerations that can be considered throughout the design of APM and provide a potential starting point for considering ethical design flows. They are a shortlist of various values and concerns that have arisen in the technology assessment literature, as well as the VSD literature for speculative technologies more specifically.[11]

1. Engage with Convergence Literature

One of the benefits of thinking of APM and other transformative technologies as being part of a converging technology landscape is that overlap of common values between the different technologies can arise, and those common values such as safety, privacy, usability, effectiveness, autonomy, etc. can serve as the basis for design.[12] Understanding APM as being co-constituted by AI, biotechnology, and information and communication technologies means a more holistic design workflow can be engaged in. This does not mean that foundational work in each technology cannot be done without reference to the other, but foundational work in each field can, and does, contribute to the others.

This convergence framing means that designers, when eliciting stakeholder values, should frame their elicitations in a way that acknowledges this co-constitutive, dynamic and changing nature of the technology in question. Not doing this can prove deleterious given the fecundity of instrumental view of technology (ie, technology is just a neutral tool) and technological determinism (ie, humans have no influence on the future development of technology). Either of those two positions leads to severe blind spots. The interactional view of technology, that on which VSD is predicated, argues for the co-constitutive nature of technology. For at least the past six decades this has been the contention of the sociology and philosophy of technology.[13]

2. Avoid Opaque Systems

As mentioned, AI in particular may prove essential to the successful development of APM systems. Perhaps the most useful systems for simulation potential APM models and outputs is deep neural networks. However, one of the issues with these systems is the tendency for them to be black-boxed and their decision making structures to be opaque to both users and engineers.

Although transparency is often construed as a value, it should be balanced with things like data privacy and security. Because of this, transparency should not be envisioned as an end-goal per se, but rather as a value that can be translated into design requirements that either support of constrain other values. To this end, the AI systems used should be able to balance these issues. Policy makers and industry leaders interested in AI enabled APM should look at the work done by the Foresight Institute, more particularly at the proposed use of AI systems such as those developed by OpenCog, or the CANDO platform designed by Christian Schafmeister.[14]

8 Till Winkler and Sarah Spiekermann, 'Twenty Years of Value Sensitive Design: A Review of Methodological Practices in VSD Projects' (2018) 21 Ethics and Information Technology 81

9 Batya Friedman et al, 'A Survey of Value Sensitive Design Methods' (2017) 11 Foundational Trends Human -Computer Interaction 2, 63–125

10 Batya Friedman and David Hendry, 'The Envisioning Cards: A Toolkit for Catalyzing Humanistic and Technical Imaginations' (30th International Conference on Human Factors in Computing Systems, Austin, 5 - 10 May, 2012) <http://chi2012.acm.org/> accessed 15 July 2019

11 Umbrello (n 6); Umbrello and Baum (n 2)

12 Steven Umbrello, Beneficial Artificial Intelligence Coordination by Means of a Value Sensitive Design Approach (2019) 3 Big Data and Cognitive Computing 1, 5

13 Winner, L., 2003. Do artifacts have politics? Technol. Futur. 109, 148–164. https://doi.org/10.2307/20024652

14 Allison Duettmann and James Lewis, *Artificial Intelligence for Nanoscale Design* (Foresight Institute 2017)

3. Aim for Proportionality

Designers should embrace a standard of proportionality and thus design APM systems that are physically limited to not manufacture explicitly deleterious substances, materials, weapons or self-replicating autonomous nano fabricators.[15] This naturally must vary amongst users (ie civilian vs military). Hence, balanced APM systems should be something that is openly promoted in order to limit over-engineering which can inadvertently open the Pandora's box of possible materials and systems that an APM system can manufacture.[16] The most notable example is perhaps the engineering of self-replicating machines. Functioning similar to biological cells (ie, closest paragon would be the ribosome), these machines would be able to replicate more of themselves. The concept, if possible, has many potential applications for deployment, particularly in space exploration and colonisation given the ability to meet weight restriction requirements. However, this type of system can also destabilize economic structures drastically and without notice given the ability to manufacture any goods at any time with marginal costs. There are many examples of both boons and cons that can and have already been conceived of. The point here is that engineers have to take into account early on and throughout the design process of the needs that the technology must meet given the stakeholders' values as well as the unintended effects that can emerge after deployment.

4. Aim for Transparency about System Security and Safety

Users of APM systems should be informed about not only the limitations of their APM systems but also the vulnerabilities of those systems. This becomes particularly relevant as nanotechnology converges with ICT, opening up a further range of converging socio-ethical issues (see point 1). Access to networks by remote means requires a minimum standard of both software and hardware security. Users of APM systems should be informed of any health threats that APM systems may pose during use such as the deleterious effects that nanoparticles and materials can have on organic cells.[17] Not only does the back-boxing of potential safety and security issues limit the social acceptance of technologies, particularly transformative ones, but they open up liability issues that may ultimately be deleterious to the potential benefits that such systems if designed well, can provide.
What this ultimately means is that transparency must be construed as explicability and understandability to the stakeholder sin question. Not only must information about how a system works be conveyed to the user, engineers, developer, etc., but it must be done so in a way that is understandable to those agents to permit effective interventions if needed.

5. Design for Accessibility

APM systems should be designed in such a way that fosters ubiquitous use given the above design flows. This is intended to limit exclusions based on socioeconomic status, thus promoting a more egalitarian use of APM systems.
This of course can be an innately difficult value to take into account given the many different groups that stakeholders can be sectioned off into, particularly when those individuals are members of more than one sub-group. Case studies in accessible computing provide examples of how computing systems can be tailored for different stakeholders.[18]

IV. Conclusions

There are already existent examples of technologies that are sensitive to the values and framing tools listed here. Industry and policy measures regarding financial technologies for example not only reveal a trend towards better regulation but also the inclusion of values like proportionality regarding the information that they gather from their users, as well as a tendency toward offering greater accessibility as a function of the technology itself. Applying this to

15 Umbrello (n 6)

16 Outlined in Umbrello and Baum (n 2)

17 ibid

18 Kristen Shinohara et al, 'Tenets for Social Accessibility: Towards Humanizing Disabled People in Design' (2018) ACM Transactions on Accessible Computing 11, 6

APM, proportional examinations should frame APM technology design. This framing helps designers to seek a balance between the potential benefits that the systems could produce against some of the risks associated with technological potential itself. In more practical terms, APM systems could include physical barriers that enable specific materials to be used and restrict the types of products to be manufactured. The benefits that arise from constraints are naturally to be weighed against the potential loss of manufacturing potential.

The FinTech world has shown that users are more willing to adopt systems if there is transparency regarding the potential hazards and vulnerabilities associated with adoption.[19] Hence, transparency, and how transparency is understood and instantiated in design is a critical factor to social acceptance. Finally, the ubiquitous adoption of APM systems may hinge on their ability to be accessed and used by any member of society regardless of socioeconomic status. Because of this, design considerations for broad spectrum use must be accounted for during early developmental and conceptualisation stages if the technology aims to be equitable and accessible.

The VSD methodology provides a principled approach to incorporating the values of stakeholders as design requirements both early on and throughout the development of a technology. The listed framing considerations provide a potentially useful first step that can be taken as they are distilled from across the converging technology discourse and provide a set of common values that are shared by the stakeholders of these different, yet ever increasingly interconnected artefacts. Policy makers and industry leaders should consider engaging with both the VSD discourse as well as its applications to other technologies to determine how to best modify its principled approach to specific design contexts.

19 Hyun-Sun Ryu, 'Understanding Benefit and Risk Framework of Fintech Adoption: Comparison of Early Adopters and Late Adopters' (51st Hawaii International Conference on System Sciences, Hawaii, 2018); See also Chris Brummer Daniel Gorfine, FinTech: Building a 21st-Century Regulator's Toolkit' (2014) Milken Institute, 5