Towards a Design Framework for Humanized AI

Luca Iandoli[a,b], Jia Shen[c]

[a] St. John's University, 8000 Utopia Parkway, 11439 New York (NY), USA
[b] University of Naples Federico II, Piazzale Tecchio 80, 80125, Naples (NA), Italy
[c] Rider University, 2083 Lawrenceville Rd, Lawrenceville, New Jersey, USA

Abstract

Negative impacts and unintended dysfunctions of AI are increasingly under scrutiny, and AI designers are often accused of not being considerate enough of the depth and consequences of the innovation they create. In this paper, we present a framework to assess the impact of AI solutions on their users, society, and environment to help designers to anticipate potential issues and create solutions that are not harmful and support individuals and society's betterment. Humanistic approaches to digital design inspire the proposed framework, including Positive and Inclusive design. The originality of the proposed approach derives from assessing the consequences of using computing solutions as they propagate across individual, social and environmental layers centered around the task the solution is supposed to perform. This paper presents an initial version of the framework and outlines the next steps for its full development.

Keywords[1]

Inclusive Design; Positive Computing; Design Thinking; Digital Design; AI Ethics.

1. Is Human-centered Design Really Human Centered?

Machine intelligence can be defined as the ability of a computer to perform data-driven inferences. One fundamental way this ability manifests itself is through algorithms that are supposed to improve users' lives by predicting events and suggesting or performing practical actions. Such algorithms need access to vast environmental, social, and behavioral data to increase their accuracy and support business models aimed at monetizing data flow in various ways. It can be shown that this combination of technical features and economic motivation create several unintended consequences in many spheres of human action, ranging from privacy threats to information overload and even dysfunctional social behavior or polarized public discourse and politics.

This paper argues that a new approach is needed in digital technology design to counteract these trends. In particular, we advocate for a new AI design paradigm to support human advancement and fulfill innovation and human growth potentials. This new paradigm is interdisciplinary, drawing from research in psychology, computer science, design, economics, and ethics, and aimed at promoting breakthrough innovations of meaning [16].

More specifically, in this paper, we claim that some of the negative impacts of technological developments in digital technologies, including AI, derive from design approaches that are not genuinely human-centered. First, the development of computing solutions has historically shied away from considering human qualities associated with psychological states, thinking modes, or behaviors that are hard to quantify. Second, and more importantly, design practices, not only in the computing field, have been characterized by an individualistic bias in which the consequences of using a design on the technical, social, and natural environment typically receive little or no attention. Because

Joint Proceedings of the ACM IUI 2021 Workshops, April 13-17, 2021, College Station, USA
EMAIL: iandolil@stjohns.edu (A.1); jiashen@rider.edu; (A.2)
ORCID: 0000-0002-9167-7564 (A.1)

of this bias, the analysis of this impact and the provisions of countermeasures to minimize negative consequences beyond the immediate users are neglected or not prioritized. This reductionist view of the human being is indeed not truly human-centered since human existence is embedded in a network of biological, affective, and social relationships.

In this paper, we draw ideas from humanistic approaches to design, such as positive computing and inclusive design, to identify guidelines that can overcome the individualistic bias and make AI-based computing solutions more considerate of their impact on individual, social, and environmental wellbeing. We show how the proposed framework can be leveraged to address explainability, fairness, transparency, and biases in AI solutions and to generate guidelines for sustainable design of intelligent computing solutions.

2. Background

2.1The individualistic bias

In their book "User Friendly" [7], the authors outline the history of how the idea of user-friendly design came to be starting from the first studies on “human error” during World War II. While introducing the necessity to shape technology around human needs and capabilities, post-war Industrial Design led to increasing standardization of taste via mass consumption and generalist advertising channeled through centralized mass-media. Partly as a reaction to increasing homologation and bourgeoise conformism, the youth revolution in the 60s, especially in its liberal version that flourished in the US starting in and around the University of Berkeley campus, advocated for individual empowerment and contributed, in the digital field, to the birth of the personal computer. The "stay hungry stay foolish" motto mentioned by Steve Jobs in his commencement speech at Stanford University in 2005 was, in fact, a citation from the much older Whole Earth Catalogue, a paper publication popular in the '60s focusing on ecology, alternative education, do it yourself, and holism, and featuring the slogan 'access to tools'.

While the techno-hippy movement gave a fundamental contribution to the digital industry's democratization, its anti-business spirit and the technology-centered mindset posed a barrier to the mainstream public formed by non-tech-savvy individuals and consumers. The commoditization of computers and the progress in the development of interactive technology between the '80s and the '00s made digital tools even more available to unskilled users. These new tools were presented as liberating and positive forces for human betterment, especially in their networked version [1]. It is not a case that human-centered design approaches such as design thinking and the idea of persuasive technologies [4] were all born or acquired popularity in those years.

Following these trends and thanks to the advent of social media and mobile computing, contemporary approaches to digital design ended up favoring a strong tendency toward ego augmentation via hyper-customized user experience.

We argue that the ubiquitous access to highly customized information is a key force behind many distortions present in the current digital sphere. Personalized information helps us to instantiate a subjective reality in which many "truths" are possible, and manipulation is easy to enact but hard to detect. The fragmentation of the political discourse, the diffusion of fake news and online misinformation, or the design of addictive digital tools can all be explained by this obsession with designing tools for ego-augmentation.

Thankfully, the digital design community is becoming increasingly aware of these adverse effects and is developing and adopting more considerate and thoughtful approaches, such as Positive Design and Inclusive Design, that we briefly introduce in the following sections.

2.2 Positive Computing

Positive computing refers to the design and development of technology to support psychological wellbeing and human potential [2]. The positive computing field grew along with the Positive Psychology paradigm proposed by psychologist Martin Seligman. He argued that psychology had restricted itself for too long to research and treat mental problems and that it was time for a focus on what makes life worth living, i.e., positive emotions and human flourishing [13]. Ten years later, Marc

Hassenzahl made an appeal to the HCI community to move beyond the "disease model" of human technology use mainly focused on removing usability problems and frustrations because "avoiding the bad experience due to a lack of instrumentality does not necessarily equate with providing a positive experience." [10] He argues that instrumentality addresses the "how" (motor) and "what" (behavior) aspects of interaction and is not concerned with the "why" (purpose) aspect of interaction on fulfilling human needs. This new level is where technology design has the potential to create profound and worthwhile experiences.

Grounded in psychological wellbeing research and multidisciplinary foundations, Calvo and Peters [2] proposed a positive computing framework identifying circumstantial conditions that influence wellbeing and determinant factors that can be cultivated to increase it. Circumstantial conditions outside positive computing's influence are excluded from their framework (e.g., personality type, intelligence levels, socioeconomic variables), while the physical and digital environment, relationships and family, and education and life-long learning are considered. Within these conditions, and reviewing existing empirical evidence, Calvo and Peters identify eight design factors frequently mediated by technology: positive emotions, motivation and engagement, self-awareness, mindfulness, resilience, gratitude, empathy, compassion, and altruism.

2.3 Inclusive Design

Approaches such as Inclusive and Universal Design [17] show that designers have not been indifferent to fairness issues. Following a definition proposed by the British Standards Institute, inclusive design is the design of mainstream products and/or services accessible to and usable by as many people as reasonably possible. Inclusive design is mindful of users with reduced ability or impairment and aims at stretching the reach of a product as much as possible beyond the ideal target. Inclusive design is fair and can help spur innovation by taking into consideration the needs of specific categories of users. For instance, while helping readers with visual impairment, audiobooks also provide opportunities for alternative book fruitions for users who do not have issues with their eyesight.

3. A Framework to Design Humanized AI

In our framework, we combine positive and inclusive design with anthropological theories on the role of technological artifacts in human and cultural development, based on the concept of artificial envelope proposed by Andre' Leroi-Gouhran [8]. By focusing on well-being, equity, and the positive impact of technology beyond functionality and the user's immediate environment, we will show that our approach can help designers to overcome some of the negative consequences of ego-augmentation in the development of computing solutions.

More specifically, the proposed approach aims at rebalancing the focus on individuals with a focus on the design's impact on the action space in which users are physically and socially embedded. We define the action space as the physical social, and ecological space where the consequences of using a design can be felt. Following research in positive computing, the action space extends beyond the immediate execution of a task across three levels:

- The Self (intrapersonal)
- The Social (interpersonal)
- The Transcendent (extra-personal)

The framework then help identifiying design factors for each level, as shown in table 1, which reports an initial and not exhaustive list of such factors.

The first level for humanized AI design is at the *self*, or *intrapersonal* level. Factors at this level are experienced within oneself, and this experience is generally not dependent on others' presence.

We include here the traditional categories of usability and ergonomics since a design must first and foremost decently execute a task in a way that is not harmful to users. However, the impact on the individual extends beyond these categories to include factors such as fluent information processing [11] and determinants of psychological well-being.

The second level is *social* or *interpersonal*, and factors at this level are dependent on the interaction between oneself and others (i.e., relatedness, empathy, trust, social status).

Table 1
Levels of human needs and factors of design

Need Levels	Design Factors	Approach
Transcendent (Extra-personal)	Altruism Compassion Ethics Environment …	Proposed Humanized Design Framework Inclusive Design Zero-waste, Circular Economy
Social (Inter-personal)	Trust Empathy Relatedness Cultural compatibility Status …	Positive Design Design Thinking, Experience Design
Self (Intra-personal)	Fluency Aesthetics Emotions Self-awareness Competence Autonomy …	HCI
	Ergonomics Usability	Usability Engineering

The last level, the *transcendent* or *extra-personal*, is characterized by factors involving thoughts or actions for the greater good and for beings and spaces beyond those we know or inhabit personally (i.e., social responsibility, compassion, ethics, and environmental concerns).

We provide a visual representation of the proposed framework in table 1.

Table 1 lists the three levels, some of the main factors considered in each level as outlined in the literature, and the design approach in which these factors have been predominantly investigated. The transcendent level still lacks design methodologies that deliberately analyze such higher-level impact, although Inclusive Design and green approaches to economy and production try to systematically address these issues. The boundaries between design approaches and levels have to be considered necessarily blurred, something that the table does not convey.

In the following, we elaborate on the framework's main levels without the ambition of being exhaustive given the vastity of the topics and the limited space available for this paper.

3.1 The Self

At the *self or intrapersonal level*, we find the three levels of processing proposed by Norman in his book on Emotional Design [9] provide a solid foundation to understand how design attributes affect individual user needs. Norman indicates visceral, behavioral, and reflective levels of processing that lead to emotions in users. Visceral responses are fast and subconscious. They are grounded on attraction and aesthetics and unrelated to product's usability and effectiveness.

We allocate **Usability** and **ergonomics** at the the foundation of the self level for two reasons. First, these factors can influence higher-level psychological constructs driven by the physical (dis)comfort of using a tool. Second, we want to safeguard the notion that a design must be first of all be functional and safe. Other key factors at the self level involve learned skills. At this level, we find Ryan and Deci's Self-Determination Theory (SDT) as most relevant [12]. SDT posits that autonomy, competence, and relatedness are the key components of motivation and wellbeing. To be self-determined, we must feel i) **autonomous**, that is, be able to attribute the outcomes of our activity to our intentions, and ii) **competent** or confident in our ability to meet challenges.

The reflective level is the home of conscious cognition. It reflects events that have happened and the source of the highest level of emotions. Previous work on emotional intelligence and Cognitive behavior therapy (CBT) supports this and indicates that **self-awareness** is the key factor at this level.

Another design factor that is critical at the self level is fluency. Fluent processing of an object is associated with positive aesthetic response and pleasure. Fluency enables the user to enter an effortless state conducive to flow and other positive emotions and impacts the user's ability to resolve complexity [11].

3.2 The Social

At the *social or interpersonal level*, SDT indicates **relatedness** (feeling secure and connected to others) as a critical factor contributing to our sense of self-determination and wellbeing. In our relationships with others, **empathy** plays an essential role, including emotion recognition, vicarious feelings, and perspective-taking [15]. Further, both cognitive empathy (the ability to recognize emotions and intentions in others) and affective empathy (ability to share feelings with others and react with appropriate emotion) are revealed [15]. Related to technology design, affective computing [5] has shown promising results in mediating strategies such as perspective-taking, emotion recognition training.

Finally, as our partnership with technology is deepening in the coming years, especially with AI, **trust** will become an increasingly important factor [7]. Not only does technology need to be designed to fit our needs, but it will also have to gain our trust. At the machine level, related to autonomy and competence, as previously discussed, trust needs to be built to put humans in control and feel capable. At the social level, for machines to be woven into our social fabrics, they need to behave in socially acceptable ways, which are often subtle. Applying Don Norman's concept of affordance, AI-specific affordance needs to be carefully designed as affordances are becoming *psychological* without physical interfaces, and the users' mental models are guided by their social expectations.

3.3 The Transcendent

Humans strive to go beyond satisfying their own immediate needs at the transcendental level and aspire to a more profound connection with the world around them.

Compassion refers to “the feeling that arises in witnessing another’s suffering, and that motivates a subsequent desire to help” [6]. While compassion describes a desire to act, **altruism** is the action. An altruistic act confers benefits on someone else at a cost to oneself. Studies show that comparison, including self-compassion and altruism, benefits well-being, including increased social connectedness, reduced stress, inspiration, and greater happiness. In recent years, studies are emerging to explore technologies designed to foster compassion and altruism, such as compassion meditation, role-play games, and design for inspiration [3]. Providing effective design guidelines at the transcendent level, Humanized AI has the potential to further these efforts to benefit human flourishing.

3.4 Example

In this brief example, we make an attempt to apply the proposed framework to a computing application: the development of data-driven algorithms to populate user feeds in social media. The arrangement of the content in our feeds depends on many factors that algorithms take into account to determine what has to be shown, when, and with what level of priority. Typically, users are not aware of how this ordering takes place, so generally speaking, these algorithms tend to score really low on metrics such as explainability, transparency, and bias. Using our framework (tab.1), we can be more systematic in assessing the impact of these algorithms' design on the users' action space.

At the functional level, personal feeds work reasonably well and quite usable. Ergonomic aspects can be evaluated based on the effectiveness and efficiency of the typical interaction modalities available on a touchscreen (or desktop) device. Continuous scrolling, distraction, and focus disruption due to notifications are areas for improvement. Still at the self level, but moving up towards the cognitive and psychological impact, the effects of navigating personal feeds on social media have a controversial impact on psychological well-being, with increasing evidence of negative effects such as information overload and anxiety via the FOMO syndrome (fear of missing out). The interpersonal and societal impact of social media is also subject to increased scrutiny. For instance, Settle [14] shows how the functioning of personal feeds in Facebook generates political polarization via exposure to "disagreeable others" and the involuntary exposure to political news and commentaries that end up being mixed with other types of information provided by friends.

Limited attention to ethical considerations in social media platforms' design has drawn increased scrutiny in recent years. The absence of filters or control does not help contain the

diffusion of dubious quality or blatantly fake information. Finally, at the environmental impact level, algorithms' design is not optimized to minimize the energy consumption and associated level of CO2 emissions produced by massive, always-on data centers. Under the increasing pressure from the public and regulators, social media and web companies are providing increasing efforts towards containing emission (Amazon), promoting AI ethics, and supporting development of intelligent filters recognizing and suppressing hate speech, cyber-bullying and racism (Facebook). Research on ethics and on the dark side of social media is equally growing [18].

A more in-depth assessment could be carried out with respect to specific indicators applied to each level, such as the above-mentioned explainability, fairness, transparency, and bias. These metrics can differ depending on the level at which these variables are assessed on the context. For instance, at the functional level, personal feeds score quite low in explainability and transparency (the platform does not provide users with an explanation of why posts are ordered in the way they are), in fairness (paid posts receive more visibility), and for bias (posts can receive more visibility and feedback because they are biased). A similar analysis can be carried out at any level for useful indicators. For instance, assessing explainability at the level of psychological well-being requires identifying satisfaction drivers that make an explanation fulfilling and convincing.

4. Conclusions

Our contribution in proposing the framework to design Humanized AI resides in identifying three levels of analysis and mapping the design factors that are most important to human wellbeing in the context of human-AI interactions at each level. The factors can be used as entry points for designers to improve AI applications in a few ways:

1. Applying the framework *actively* in designing new AI applications dedicated to promoting wellbeing explicitly
2. Preventatively, to address or prevent detriments to wellbeing in ongoing AI projects
3. Retrospectively, in evaluating and improving existing AI applications

In the next steps of our work, we aim to fully specify the framework in terms of design factors and metrics and apply it to the assessment of specific AI applications.

5. References

1. Y. Benkler, The wealth of networks: How social production transforms markets and freedom. Yale University Press, Princeton (NJ), 2006.
2. R. Calvo and D. Peters, Positive Computing, MIT Press, Cambridge (MA), 2017.
3. E. Fehr ad U. Fischbacher, The nature of human altruism. Nature 425 (2003) 785-791.
4. B.J. Fogg, Persuasive technology: using computers to change what we think and do, Morgan Kaufmann, San Francisco (CA), 2003.
5. K. Gerdes, E. Segal, and C. Lietz, Conceptualizing and measuring empathy, British Journal of Social Work 40 (2010) 2326-2343.
6. J.L. Goetz, D. Keltner, and E. Simon-Thomas, Compassion: An evolutionary analysis and empirical review, Psychological Bulletin 136 (2010) 351-374.
7. C. Kuang and R. Fabricant, R., User Friendly: How the Hidden Rules of Design are Changing the Way We Live, Work & Play, MCD, New York (NY), 2019.
8. A. Leroi-Gourhan, Le geste et la parole. Technique et language, Albin Michel, Paris (France), 1964.
9. D. A. Norman, Emotional Design: Why We Love (Or Hate) Everyday Things, Basic Books, New York (NY), 2004.
10. M. Hassenzahl, Experience Design: Technology for All the Right Reasons, Morgan & Claypool Publishers, San Rafael (CA), 2010.
11. R. Reber, N. Schwarz and P. Winkielman, Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing

experience? Personality and social psychology review 8 (2004) 364-382.
12. R. M. Ryan and E. L. Deci, Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 55 (2000) 68-78.
13. M.E.P. Seligman and M. Csikszentmihalyi, M., Positive psychology: An introduction. American Psychologist 55 (2000) 5-14.
14. J.E. Settle, Frenemies: How social media polarizes America. Cambridge University Press, Cambridge (UK), 2018.
15. T. Singer, The Neuronal Basis and Ontogeny of Empathy and Mind Reading: Review of Literature and Implications for Future Research. Neuroscience & Biobehavioral Reviews 30 (2006) 855-863.
16. R. Verganti, Design, Meanings, and Radical Innovation: A Metamodel and a Research Agenda, The Journal of Product Innovation Management 25 (2008) 436-456.
17. R. Coleman and C. Lebbon. Inclusive design. Helen Hamlyn Research Centre, Royal College of Art, London (UK), 1999.
18. J. Hemsley, J. Jacobson, A. Gruzd, and P. Mai (eds.), Special Issue: Social media for social good or evil, Volume 4(3), Social Media+ Society, 2018.