Ref. Ares(2020)2314744 - 30/04/2020

Dear EU Commission,

As part of my doctoral training, I investigated trust design in highly autonomous systems. This research has generated in excess of eight publications related to the topic at hand.

In the following table I provide a break through of the issues presented by the white paper and how my publications addresses them.

PP	PROBLEMS	ISSUE	DEFINITION	PUBLICATIONS ADDRESSING THE ISSUE
9	**TRUST**	ASYMETRIES	UNINTENDED EFFECTS	**PROSPECTIVE DESIGN (1 PAPER)**
10	**HARM**	MATERIAL INMATERIAL	DEATH LOSS	**CASES/SURVEYS - autonomy, reparation and accountability papers** **CASES/SURVEYS - autonomy, reparation and accountability papers**
12	**PROBLEMS**	UNCERTAINTY	BLACK-BOX COMPLEXITY UNPREDICTABILITY AUTONOMOUS BEHAVIOUR	**PROSPECTIVE DESIGN (1 PAPER)** **SYNTHETIC CONSEQUENTIAL REASONING (1 PAPER)** **LEVELS OF ACCOUNTABILITY/REPARATION (2 PAPERS)** **DIGITAL RIGHT (1 PAPER)**
13	**CONTEXTS**			**PROSPECTIVE DESIGN (1 PAPER)**
14	**LIMITATIONS**	EXPONENTIALITY	FROM PRODUCTS TO SERVICES FROM LAUNCH TO UPDATES FROM MAKER TO 3rd PARTIES	**LEVELS - autonomy, reparation and accountability papers** **LEVELS - autonomy, reparation and accountability papers** **LEVELS - autonomy, reparation and accountability papers**
15	**LIABILITY**	PROTECTION		**LEVELS - reparation and accountability papers** **DIGITAL RIGHT - the right to reparations paper**
17	**DEVELOPMENT**	RISK	LOW HIGH	——— **RISK MATRIX AND CALCULATION TOOL - optimisation paper**
19	**DESIGN**	DATA TRANSPARENCY ALGORITHM	DIVERSITY AND PRIVACY INTENTIONS ROBUSTNESS/ACCURACY/RELIABILITY	**THIS AREA WENT BEYOND THE SCOPE OF MY ENQUIRY** **LEVELS - autonomy, reparation and accountability papers** **LEVELS - autonomy, reparation and accountability papers**
21	**OUTPUTS**	HUMAN CONSENT		**LEVELS - autonomy, reparation and accountability papers**
22	**ADDRESSES**	ACTORS		**LEVELS - accountability paper**
24	**COMPLIACE**	INTERVENTIONS	EX ANTE EX POST	**SIMULATION (1 PAPER) - optimisation paper** **CALIBRATION- Autonomy and synthetic consequential reasoning papers** **REPARATION - reparation and accountability papers**

In this process four publications were submitted and were accepted by the National Data Strategy Board. The committee accepted all publications as evidence to inform the development of the framework which will determine the use of Artificial intelligence in the UK. (See papers 1 to 4)

1. Galdon, F., & Wang, S. J. (2019a). Designing trust in highly automated virtual assistants: A taxonomy of levels of autonomy. Artificial Intelligence in Industry 4.0: A collection of innovative research case-studies. International Conference on Industry 4.0 and Artificial Intelligence Technologies IAIT. Cambridge, UK.
2. Galdon, F., & Wang, S. J. (2019b). From apology to compensation; A multi-level taxonomy of trust reparation for highly automated virtual assistants.
3. Galdon, F., & Wang, S. J. (2019c). Addressing accountability in highly autonomous virtual assistants. Proceedings of the 1st International Conference on Human Interaction and Emerging Technologies (IHIET 2019) conference August 22-24, 2019, Nice, France.
4. Galdon, F., & Wang, S. J. (2019d). Optimizing user engagement in highly automated virtual assistants to improve energy management and consumption. Proceedings of the Applied Energy Symposium, MIT Media Lab, 22-24 May 2019.

5. Galdon, F., Hall, A. (2020b). Synthetic Consequential Reasoning: Facilitating the design of synthetic morality in highly automated systems via a multidimensional-scalar framework. Proceedings of the 2nd International Conference on Human Interaction and Emerging Technologies: Future Applications (IHIET-AI 2020) Lausanne, Switzerland.
6. Galdon, F., Hall, A., & Ferrarello, L. (2020c). Designing trust in Artificial Intelligence: A comparative study among specifications, principles and levels of control. Proceedings of the 2nd International Conference on Human Interaction and Emerging Technologies: Future Applications (IHIET-AI 2020) Lausanne, Switzerland.

7. Galdon, F., Hall, A. (2020a). The right to reparations: a new digital right for repairing trust in the emerging era of highly autonomous systems. Proceedings of the 2nd International Conference on Human Interaction and Emerging Technologies: Future Applications (IHIET-AI 2020) Lausanne, Switzerland.

8. Galdon, F., Hall, A. & Wang, S. J. (2019f). Prospective design: A future-led mixed-methodology to mitigate unintended consequences. Proceedings of the International Association of Societies of Design Research Conference IASDR2019, The University of Manchester, UK.

The first paper represents the foundation of the research in trust design in automation. This paper presents the first ever levels of autonomy in the contexts of Highly Automated Virtual Assistants.

Paper two and three present two fundamental variables; reparation and accountability.

The combination of these three papers forms a multi-dimensional framework to address the impossibility of monitoring/controlling complex dynamic systems and propose reparation and accountability as strategies to build and maintain trust in these systems. This approach knowledge that trust in automation starts before the interaction and follows thereafter.

Paper four builds from this framework to provide a method of calculation to generate a trust rating by which this score can be used to optimize users‘ engagement. (in this case via a case study on energy management and consumption). This tool aims to provide a tool to simulate potential interactions impact.

Papers five and six present two evaluative exercises on the Trust calculator. These evaluative exercises aimed to understand whether this tool could be used for designing highly automated systems from an integrative multi-stakeholder perspective.

Paper seven acknowledges that we cannot fully monitor emerging systems due to its increasing complexity. Therefore, this paper presents a new digital right to ensure that emerging Highly Automated Complex Systems (HACS) interactions remain accountable while the development of these technologies cannot fully guarantee its behavior.

Finally, with the advent of the digital age, accelerating technology complexity, black box technologies and wicked problems new prospective approaches in design research are required to deal with the exponential nature of our emerging digital era. In this context, paper eight present a guiding design methodology to mitigate unintended consequences.

Please do not hesitate to contact me if any question arises.

Kind regards,

Fernando Galdon
Ph.D. Candidate, Global Innovation Design at the Royal College of Art, London
Visiting Tutor, Innovation Design Engineering at the Royal College of Art, London