ESDIT RESEARCH STRATEGY 2024-2029

1. Introduction

This document builds on and partially replaces the March 2022 ESDiT Research Strategy, which is the most recent document outlining a research strategy on ESDiT. The March 2022 document was developed to have a more detailed research strategy based on the ESDiT research proposal that was submitted in 2018. The document is based on several consultation rounds with the Management Board, consortium and coordinators.

The aim of this document is to define strategic research actions to be taken in the period 2024-2029 to best achieve the research objectives at the end of the programme. This is a high-level strategy, and – at least in the current version - the proposed strategic research actions will not make reference to proposals for specific projects, nor is there a determination of the number of needed new projects per objective.

2. ESDiT's research objectives

The overall goal of ESDiT is the following:

> ***In ESDiT, we aim to innovate the ethics of technology so that we can critically evaluate and guide the development, introduction and use of current and future socially disruptive technologies.***

We achieve this goal through five objectives, expressed in the following five research questions:

1. *Understanding the disruptive effects of 21st century SDTs.* What are the social, political, philosophical and ethical implications of the new generation of SDTs in the 21st century, and what are the socially disruptive impacts that they have on humans, nature, and society, particularly new digital, bio and brain and environmental and sustainable technologies?

2. *SDTs and conceptual disruption.* What are the key philosophical and ethical concepts that are challenged by SDTs, and what reassessments, revisions, and innovations are needed in response, taking into account philosophical insights from non-Western traditions?

3. *New approaches for ethical assessment and guidance of SDTs.* How can we develop new, comprehensive, and inclusive approaches in ethics and philosophy for analyzing, morally evaluating, guiding and intervening in the development and implementation of socially disruptive technologies, specifically the newest generation?

4. *Technology ethics and multi- and transdisciplinarity.* How can we innovate the ethics and philosophy of technology by developing new collaborative approaches between philosophers, engineers, social scientists, policy makers, designers, and artists, aimed at improved philosophical and ethical analysis and responsible innovation?

5. *Innovating practical philosophy.* How can revisions of philosophical concepts and other results serve to innovate the field of practical philosophy?

The relation between these five objectives is as follows. Objectives 2 through 5 contribute to the aim of innovating the ethics of technology. They do so in complementary ways. Objective 2 does so by developing approaches for the understanding, analysis and assessment of conceptual disruption by SDTs. Objective 3 does so by developing general approaches for the ethical assessment of SDTs, which will consider any type of social or conceptual disruption engendered by these technologies. Objective 4 contributes through the development of new models or approaches of multi- and transdisciplinarity, which are needed to advance the field of ethics of technology. Objective 5 does not contribute to the innovation of ethics of technology specifically, but to the field of practical philosophy as a whole, of which ethics of technology is a part. Finally, objective 1 is not concerned with innovating the field, but rather with analyzing and assessing 21st century SDTs and their consequences. This objective depends on the other four objectives since the innovations generated in them support better analysis in objective 1.

While objective 5 logically comes a bit later than objectives 1 through 4, we are not following a waterfall approach, and we want research for the different objectives to influence each other.

We will now turn to our proposals for research for these objectives for the period 2024-2029.

3. The *Conceptual disruption* objective

For this objective, significant progress has been made in developing theories and methods of conceptual disruption and conceptual change. A start has been made for research on the role of SDTs in conceptual disruption and on conceptual engineering. Both topics need to be studied more. The results of case studies on specific technologies and concepts are currently not well-documented. The impression is that not all these projects have paid much attention to conceptual disruption, and that few make use of the theories and methods developed in the F&S line. Nevertheless, there should be some interesting analyses in the case studies.

It is proposed that future projects do not focus on particular technologies and their disruption of particular concepts but focus on particular concepts or conceptual clusters (probably the latter).

These are then studied in the context of multiple SDTs simultaneously. In the selection of concepts/conceptual clusters, their societal relevance and relevance to current debate should be taken into account. Past histories of conceptual change should also be taken into account. We add that if it does not interfere with this strategy, there can still be some new studies that focus on particular technologies, particularly technologies that were not studied so far, or interesting conceptual disruptions associated with technologies we have already studied.

More future emphasis will be put on comparative analysis, synthesis, proposals of new concepts/conceptual frameworks, and integration of results of the intercultural line. We also consider how concepts/clusters have changed over time in history. To know when and how to develop improved and new concepts, we must also focus on conceptual resilience in the light of disruption and develop assessment tools to determine under what conditions conceptual engineering is called for. Closer collaboration is needed between researchers working on the conceptual disruption objective. We need an organization structure that supports such close collaboration.

4. The *New approaches* objective

Not many projects have been started so far with this objective. Björn Lundgren was appointed specifically to study and propose methods in ethics of technology. He has made some interesting contributions, including a method for addressing contextual challenges to ethical guidelines, but his project had an early ending. Recently, Haizea Escribano-Asensio has started a project aimed at developing new approaches for taking into account technological and social structures in ethics. Sergio Urueña-Lopez (funded externally to ESDiT, but fully participating) has studied the role of foresight in ethics of technology, and has developed a framework for assessing the disruptive affordances of anticipation.

In the future, we do not want to focus on methodological and theoretical innovation in general, but specifically on new theories and methods for assessing and guiding SDTs. These theories and methods should build on the results from research on the conceptual disruption objective and the 21st century SDTs objective. The focus will be on approaches to ethical analysis of SDTs, since approaches for ethical guidance are covered in the multidisciplinarity objective. However, some approaches for ethical guidance could still be developed in the context of this objective if they are an outgrowth of the work on ethical analysis. They are then used to inform the work on guidance that takes place for the objective on multidisciplinarity.

Projects on approaches for the political philosophy of SDTs could also be included with this objective (in relation to assessment and guidance). New approaches developed in the intercultural philosophy track should also be used to feed into this objective.

Most research for this objective will take place at a macro- and meso-level, whereas most in the multidisciplinarity objective will take place at a micro-level. Since both objectives focus on

methods and approaches, there should be moments of exchange and collaboration between the two objectives.

5. The M*ultidisciplinarity* objective

Past research for this objective has taken place in the context of the STEM and Art tracks, and in Elisa's Paiusco's project, which considers the role of policy. Research so far has been exploratory, with not many results yet.

We propose to understand this objective in terms of *guidance* of (socially disruptive) technologies. Guidance is here understood broadly, also including for example critical reflection and deliberation about (socially disruptive) technologies (as in the art track). This guidance focuses primarily on the micro and meso level, e.g. concrete design projects or the company level. Normative appraisal of socially disruptive technologies and normative reflections on the institutions needed to properly deal with socially disruptive technologies (political philosophy), we see as primarily falling under other research objectives (3 and 1) (although concrete projects can of course contribute to more than one objective).

Guidance of technology requires a multi- and transdisciplinary approach, and therefore this objective is also aimed at facilitating and exploring new forms of multi- and transdisciplinary cooperation (as emphasized for example in the original description of the STEM track). Such guidance is (almost by definition) intended to have also social impact, and may include recommendations to designers, companies and policy makers.

For the second period of the ESDiT project, we propose to focus on:

1. Developing new *approaches and methods* for guidance of (socially disruptive) technologies, especially for engineering design and for the interaction between ethics and art

2. *Complementarity* with other research programs and initiatives; collect e.g. outcomes of these and try to capitalize on their outcomes; This includes (but is not necessarily limited to):

    a. Other Gravitation projects, in particular HI (Hybrid Intelligence)

    b. ELSA Labs

    c. Existing collaborations with e.g., RSL and De Waag

6. The *21st century SDTs* objective

After the focus on PhD projects on various SDT's – AI, Social Media, geoengineering, synbio etc. - involving multiple concepts - democracy; normative uncertainty; nature etc. – and the

development of research on climate technology in the first half of the program, the second half of the program will strategically focus on the synthesis of existing results in current PhD/post-doc projects in order to identify general patterns of SDTs and social disruptions in the 21st century that can inform theory and methods development. We are interested in general patterns emerging from SDTs – typology of technologies, categories of disruption & impact, disrupted concepts involved etc. – and in general domains of impact and patterns of interaction – human, nature, society – and the social disruption of the lifeworld.

The objective is not to develop one unifying theory, but to explore various theoretical angles and conceptualizations. There will be room for new case studies of specific SDTs, but preferably in the context of synthesizing theories and only in case these SDT are not covered yet and expected to have high impact (for example, SDTs (AI) in animal breeding and husbandry, gene therapy in health).

7. The Innovating practical philosophy objective

We were planning most of the work for this objective to be realized later on in the programme, when we have advanced significantly with the conceptual disruption objective. While it is not too early to engage more with practical philosophers outside our field, it may be too early in many cases to present these results. This should be considered on a case-by-case basis. When we have more results to share, we will engage progressively more with "mainstream" practical philosophy.

In the coming years, we should invest in relationships and in becoming more visible in practical philosophy. We should strive to become a serious conversation partner. Perhaps there should be a taskforce that supports people in doing so. We will organize joint workshops with "mainstream" philosophers and have special issues in appropriate journals. We should also team up with other applied fields (environmental ethics, bioethics, etc.) in order to use insights from applied ethics fields to feed into theories and methods of practical philosophy.

Much of the research for this objective could take place as part of projects that also have objectives 1, 2 or 3 (and maybe 4) as an objective. There could also be one or two projects that specifically focus on implications of our research for practical philosophy, but these should then probably take place near the end of the project.

The results that we want to use to impact practical philosophy will probably mainly come from objective 2 (conceptual disruption; especially our proposals for new concepts) and 3 (new approaches). Part of the investigation will be how classical practical philosophy will have to change in order to accommodate our results, so that they also work for SDTs and their ethical and philosophical challenges.