Harnessing robot experimentation to optimise the regulatory framing of emerging robot technologies

Carlos J. Calleja*

Hadassah Drukarch*

Eduard Fosch-Villaronga*

* *eLaw Center for Law and Digital Technologies at Leiden University*, *e.fosch.villaronga@law.leidenuniv.nl*

Abstract

Robotics and AI are dynamically and rapidly evolving, introducing the problem of regulatory disconnection, where either "the covering descriptions employed by the regulation no longer correspond to the technology" or "the technology and its applications raise doubts as to the value compact that underlies the regulatory scheme" (Brownsword, 2008). Lack of information seems to be one of the main drivers behind such a disconnection. Equipping regulators with better means to understand and tackle novel technologies thus demands solving information asymmetries among developers and policymakers. This paper introduces PROPELLING: 'Pushing robot development for lawmaking,' an FSTP (financial support to third parties) from the H2020 EUROBENCH project— as an alternative for remediating those asymmetries. Using ISO's safety standards for lower-limb exoskeletons as a case study, PROPELLING probes testbeds as data generators for standard-makers. Its central tenet is that testbeds offer a promising setting for bringing policymakers closer to research and development (R&D), and, in this way, they can be harnessed to unravel an optimal regulatory framework for emerging technologies that is based on science and evidence. The piece is structured in four sections. Section 2 argues that the paucity of data encumbers defining adequate policies for robotics. Section 3 introduces PROPELLING as a model for addressing those difficulties, whereas Section 4 highlights the significance of experimentation for improving the content of standards. Section 5 presents some of the difficulties of experimentation and early lessons learned. It also argues for the importance of testbeds and data repositories as a proper setting for replicable experimentation and information gathering.

***Keywords* –** policymaking; evidence-based; robots; regulation; safety; standard-setting; testbeds; exoskeletons.

1 Introduction

Robotics and AI growingly feature in healthcare contexts due to their increased roles and capacities in performing surgery, helping in rehabilitation or therapy, mainly as an upshot of the need to reduce human contact (Alemzadeh et al., 2016; Aymerich-Franch & Ferrer, 2020; Fosch-Villaronga & Drukarch, 2021). Advances in healthcare robotics and AI may entail incredible progress for medicine and healthcare delivery soon and could eventually help repair misdiagnoses and their very high consequences for society (Singh, Meyer, & Thomas, 2014). Still, inserting robots in such remarkably sensitive domains raises puzzling legal and ethical considerations (Vallor, 2011; Wynsberghe, 2013; Palmerini et al., 2016; Fosch-Villaronga, 2019). As with other emerging technologies, this field is dynamically and rapidly evolving, primarily due to its revolutionizing capabilities in increasing productivity and resource efficiency. However, its rising autonomy, fastness, increased roles and capabilities, and novelty are questioning existing regulations' fitness (Leenes et al., 2014; Fosch-Villaronga et al., 2021). Indeed, technology is capable of leaving the law behind at any phase of the regulatory cycle, pointing to the problem of regulatory disconnection, where either "the covering descriptions employed by the regulation no longer correspond to the technology" or "the technology and its applications raise doubts as to the value compact that underlies the regulatory scheme" (Brownsword, 2008; Brownsword & Goodwin, 2012).

Information disparity seems to be one of the main drivers behind such a disconnection. As science moves faster than moral understanding, people even struggle to articulate their unease with the perils novel technologies introduce (Sandel, 2007). Furthermore, it is common to see inventors and users sidelining ethical considerations while focusing on the practical considerations of efficiency and usability (Carr, 2011).

Regulation is not immune to those problems. On the contrary, information asymmetries between corporations and regulatory agencies are increasing, impeding the enactment of frameworks that are closer to reality and more attuned with the real problems that technology poses. The recent work conducted as part of the LIAISON Project[1], a subproject from the H2020 COVR project, indicates that less than a third of the involved industry experts believe that there is no link between robot development and policy or standard making (Fosch-Villaronga and Drukarch, 2021). 66% believe that such a link does exist but that this link is either far too complex and lacks openness, or only exists between robot development and policy/standards making. Only a tiny 7% believes that such a link exists between robot development and policy/standard making.

This all points towards the idea that regulators often operate in a regulatory environment where it is difficult for them to enter into the conversation, let alone intervene adequately. Should these asymmetries continue, technology companies "[will] have a lock on how their products work while underfunded and understaffed regulators will continue to struggle not only to understand the technology but to articulate their concerns" (Guihot and Moses, 2020). Moreover, (robot) developers will continue to struggle to implement legal provisions into their designs, resulting in continuous disconnects between policy goals and safe technology (Kapeller et al., 2021).

Equipping regulators with better means to understand and tackle novel technologies thus demands solving information asymmetries among developers and policymakers. However, the key question is how and what kind of information can bridge this gap. How to ground regulatory interventions on accurate information about technologies, their risks, and adequate safeguards?

This paper introduces PROPELLING: 'Pushing robot development for lawmaking,' an FSTP (financial support to third parties) from the H2020 EUROBENCH project— as an alternative for remediating those asymmetries. Using ISO's safety standards for lower-limb exoskeletons as a case study, PROPELLING probes testbeds as data generators for standard-makers. Its central tenet is that testbeds offer a promising setting for bringing policymakers closer to research and development (R&D), and, in this way, they can be harnessed to unravel an optimal regulatory framework for emerging technologies. That is, one based on science and evidence.

This paper introduces some of the difficulties in regulating emerging robots in section 2. It focuses on how the lack of data encumbers defining adequate interventions. Section 3 introduces PROPELLING as a model for addressing those difficulties and serving as a means to generate an optimal regulatory framework for emerging technologies. Section 4 focuses on ISO 13482:2014 (Robots and robotic devices — Safety requirements for personal care robots) to highlight the significance of experimentation for improving the standard's content. Section 5 presents some of the difficulties of experimentation and early lessons learned, along with the importance of testbeds and data repositories as a proper setting for replicable experimentation and information gathering. The paper concludes with a summary and a reflection on how science could support policies framing robot development.

2 The conundrum of regulating emerging robots: lack of information, general principles codes, and the quest for better norms

A review of the literature reveals a paucity in harnessing R&D outcomes to improve existing regulatory instruments (Fosch-Villaronga and Heldeweg, 2018). Indeed, while the chemical, food, and pharmaceutical industries established years ago use evidence-based models that ensure the safety of these products EU-wide, these frameworks have yet to be seen within robotics regulation.

Consider the role of evidence in influencing food and nutrition-related public health policy. Within this field, for instance, the United Nations' Food and Agriculture Organization (FAO) has been developing methods, data repositories and training programs to strengthen countries capabilities to support interventions on credible data and statistics. The case of Mexico's tax on sugar-sweetened beverages is an example in point. Here, the correlation between increases in sugar consumption and the rise of obesity during the previous decades, informed the adoption of an excise tax on sugar-sweetened beverages (Rocha & Harris, 2019)

Albeit there is no model easily transposable to the realm of robotics, the contrast serves to highlight how the regulatory landscape for these devices is currently populated by a myriad of technology-neutral regulations (Leenes et al., 2017), abstract codes of conduct, and trustworthy-based

1 *See* https://www.universiteitleiden.nl/en/research/research-projects/law/liaison.

ethical guidelines (HLEG AI, 2019). These lack the necessary empirical grounding to inform researchers, designers, and developers' practices adequately. Moreover, they fail to tackle the spinous problem of validating the introduction of design measures supposed to make robots fit those abstract values (Valori et al., 2021; Cf. Lipton, 2016). Not rarely, developers find themselves with unclear regulatory guidance. In this scenario, they are often left to decide if and if so, how the development process can be continued: a) positive fit, valorize; b) unclear legal fit, ask clarification, permission, assume negative; c) negative/no fit, stop, adapt, lobby, ignore (Fosch-Villaronga and Heldeweg, 2018).

Lack of adequate information is among the reasons behind the scarcity of adequate interventions. How much regulators know about technology, its effects, or the regulatory environment is critical for regulating emerging developments. Having this feature in mind, Koops distinguishes between known unknowns, i.e., we know the technology and that it has some impacts, but do not know what those impacts are; unknown unknowns, i.e., where even the existence of the impact is ignored (2010).

Emerging robotics, and particularly lower-limb exoskeletons, are prone to both types of phenomena. These body-borne devices are "inextricably intertwined with the human body" (Mann, 2012), and, as such, they raise particular questions in different dimensions that range from safety, to data protection, dignity, agency, control, and trust (Kapeller et al., 2020). However, the content of those impacts remains largely under-theorized (ibid). Think of the impact of data security on user's safety. Exoskeletons are cyber-physical systems directly fastened to the user's body (Fosch-Villaronga et al., 2018; Greenbaum, 2015b) that can have vulnerabilities in security which can compromise the correct functioning of the device. While a myriad of safety hazards can arise from cybersecurity vulnerabilities in cyber-physical systems, the extent of that impact is still to be entirely determined (Morante et al., 2015; Fosch-Villaronga & Mahler, 2021).

Beyond specific impacts, the methods for assessing risks also remain a contested matter. He and colleagues (2017) documented the lack of uniformity on inclusion and exclusion criteria for testing subjects, along with the reports of adverse events and identified risks. This lack of evaluation methods and reporting systems opens up a gulf between safe devices and those whose hazards remain unknown due to poor framing and testing (Tucker et al., 2015). Furthermore, due to their close interaction with users, these devices that have traditionally been assessed against physical safety requirements introduce a cognitive dimension that involves psychological aspects from the user (Martinetti et al., 2021). Some of them can relate to shared control, visual appearance, or fear of falling. However, the nature of these impacts and how they can be appraised remains unexplored mainly (Fosch-Villaronga, 2019). Hence, there is room for unknown unknowns, where we are yet to be even aware of the existence of potential hazards (for some few remarkable exceptions, *see* Kapeller et al., 2020; Fosch-Villaronga, 2019; Pons 2010).

Estimating risks and choosing proper safeguards is thus a central challenge for regulating exoskeletons —and certainly other emerging technologies. Crafting policies on the verge of further development demands, on the one hand, dealing with scant to nonexistent data on hazards, and, on the other, predicting how technology will evolve, including its impacts. When it comes to robots, the problem goes beyond insufficient information, as even the test methods have not been put into action nor broadly evaluated. This is particularly relevant because, despite the advances and many clear benefits that medical robots may bring for society, systems that exercise direct control over the physical world have the potential to cause harm in a way that humans cannot necessarily anticipate, control, or rectify (Amodei et al., 2016). If responsive policymakers wished to regulate robots based on empirical data, they would undoubtedly be puzzled by the sort of information required, whether the sources are generalizable, or whether data from isolated experiments could provide the basis for governing real-life risks.

Tackling those challenges requires finding a proper setting for generating and curating the information regulators need to craft their interventions. We tried to advance in this direction in PROPELLING, a project that aims at piloting robotic testbeds as information sources for policy purposes. Bringing regulators and developers closer in an environment where robot devices and testing methods are developed and validated offers a promising path to solve the information conundrum of emerging technologies, which has already been tested in Japan for robot technologies (Weng et al., 2015).

3 PROPELLING: the model

Assessing risks through experimentation is essential to ensure robot safety and compliance with existing norms. The anticipation of hazards and reflections on appropriate safeguards often happen in testing beds, where prototypes are improved to meet safety standards. In this scenario, benchmarking proves to be a vital instrument to assess the readiness of technology. Combined with testbeds, they provide indicators of whether a specific solution is suitable and safe.

With that in mind, the European Union’s Horizon 2020 is funding EUROBENCH, a program that aims at setting the first framework for the development of benchmarks for robotic systems. Its purpose is to develop methods and tools to assess devices in a rigorous but also replicable way.[2] Interesting as they are, these reflections are mainly restricted to the development of prototypes in light of their further move into real-life applications. PROPELLING departs from the premise that they could also provide knowledge to improve regulations by, among others, establishing new safety requirements or reformulating existing criteria.

PROPELLING thus gathers a multidisciplinary team to execute EUROBENCH's protocols. The goal is to improve not only a specific device but also the overall regulation for regulating robotics. In this sense, PROPELLING will use the EUROBENCH test facilities, including software and databases, to assess a combination of indicators and control algorithms. Reflecting on the possibility to harness experimentation as a source of evidence-based knowledge for policy interventions is thus PROPELLING's objective.

The project's primary focus is lower-limb exoskeletons — specifically, Technaid's Exo-H3—, in light of ISO 13482:2014, the first safety standard devoted to personal care robots. It thus tests particular safety requirements for systematically appraising safety assessments and regulations for marketed lower-limb robotic exoskeletons. Instead of focusing on the device, however, it seeks to understand if ISO 13482:2014 addresses safety sufficiently and comprehensively. In addition, some of the safety requirements revolve around topics identified in the context of the H2020 Cost Action 16116 on Wearable Robots,[3] such as those concerning psychological aspects (e.g., fear of falling), push recovery algorithms, and different user categories (see Kapeller et al., 2020; 2021).

There are indeed some indications that ISO 13482:2014 fails to address safety comprehensively and accurately. For instance, while balance loss is the second cause of falls among older adults, travel instability measures do not apply to lower-limb exoskeletons (ISO 13482:2014 Annex A.1 Hazard item 59). Another example is that, while obstacles can pose a risk to normal gait (e.g., stairs, objects), ISO 13482:2014, however, states that collisions with safety-related objects, other robots, fragile safety-related objects, walls, permanent/unmovable barriers "are not applicable to restraint-type physical assistant robots" (ISO 13482:2014 Table 7, p. 45).

Among the myriad of hazards and safety measures, PROPELLING focuses on the fear of falling (FoF), control algorithms for protective stops and graceful collapsing, and instability in a collision (*see* next section). These elements have in common that they remain unaddressed within the ISO 13482:2014 as specific hazards or safety measures. In this sense, they serve to appraise the possibilities of harnessing information from experimentation and R&D to improve existing standards. Furthermore, PROPELLING experiments with a comprehensive set of volunteers (see figure 1). Indeed, ISO 13482:2014 recognizes that more specific standards demand more comprehensive numeric data on different categories of people (Introduction, vi). Therefore, the project aims at using test benches to understand how safety requirements apply to various categories of users.

Figure 1. PROPELLING volunteers

User	Characteristics
User 1	female, healthy, young adult (<35), weight: 65-80 kg, height: 165-180 cm
User 2	male, healthy, young adult (<35), weight: 70-90 kg, height: 170-190 cm
User 3	male, healthy, older adult (>60), weight: 70-90 kg, height: 170-190 cm
User 4	female, healthy, older adult (>60), weight: 65-80 kg, height: 165-180 cm

PROPELLING's expected outcome is thus to use evidence-based knowledge to put forward specific, informed recommendations to revise ISO 13482:2014 for lower-limb exoskeletons. This is particularly relevant because the standard has remained virtually untouched in the latest revision in 2020, while the need for more specific requirements and numeric data for different categories of people was already identified back in 2014 (ISO 13482:2014 Introduction, vi).

4 Generating evidence-based knowledge on ISO 13482:2014: the case of exoskeletons

Crafting policies on the verge of further development demands, on the one hand, dealing with scant to nonexistent data on hazards and, on the other, a great deal of prediction. When it comes to robots, the problem goes beyond insufficient information. It could be that specific hazards may be neglected because "at the time of publication, the test methods (...) have not been implemented or evaluated broadly" (ISO/TR 23482-1:2020 Application of ISO 13482:2014 safety-related test methods). It could also be that test methods and benchmarks used as part of their development have not been developed (e.g., methods for psychological risks) and multidisciplinary efforts to

[2] *See* https://eurobench2020.eu/.

[3] *See* https://wearablerobots.eu/.

incorporate ‘ELS’ (ethical, legal, and societal aspects’) have only emerged recently (Kapeller et al., 2021).

Although PROPELLING has aimed to unearth such information through extensive literature review, direct interaction with the broader community of stakeholders could be explored to gain a more in-depth insight into their knowledge in this regard. Moreover, obtaining insights into the inconsistencies, dissonances, and inaccuracies within existing regulatory efforts and initiatives towards governing robot technologies as experienced by robot developers proves to be complicated. In their initial findings, the LIAISON project indicates that 'while the data obtained through the applied methods have proved to provide valuable insight into the usefulness of LIAISON, further possibilities should be explored regarding outreach and effective stakeholder involvement in practice.' (Fosch-Villaronga and Drukarch, 2021).[4]

Interestingly, while surveys are very clear, compact, and user-friendly, despite outreach having been sought through various means and platforms, the response rate to these surveys remained very limited. The LIAISON project organized interactive workshops, community engagement, and formal meetings to build and engage with the community, avoid having a low response rate, and increase the responses' focus. Dedicated workshops and presentations were organized at the European Robotics Forum 2021 and the European Commission workshops on "Trends and Developments in Artificial Intelligence: Standards Landscaping and Gap Analysis on the Safety of Autonomous Robots," a set of dedicated webinars for the Digital Innovation Hub (DIH) communities related to healthcare and agriculture, and several formal engagements with the respective DIH communities, the industry, academia, policymakers, and industry associations. While actively and interactively involving stakeholders through these means has proven to be a valuable and appreciated format for engagement with t
he broader community of stakeholders, it highlights the difficulties in obtaining sufficient information to craft new policies to frame the field adequately.

Wearable robots (WRs) include hardware —actuators and sensors— along with control algorithms and their main goal is to augment, train or supplement human motor functions (Greenbaum, 2015a). WRs are rapidly expanding (Bergamasco and Herr, 2016), and the exoskeleton market size has been projected to reach USD 4.2 billion by 2027, growing at a compound annual growth rate of 26.3% until 2027 (Grand View Research, 2020). Still, the type and extent of potential risks of these devices are yet to be understood, and industry standards are yet to be developed (He et al., 2017). For instance, the current understanding is that while these devices have been developed using user-centered methods, they have revolved around improving the patient’s conditions without thinking about how WRs would impact caregivers’ work. As a result, caregivers are continually seen in strange, uncomfortable, and painful positions to help WR users perform rehabilitation.

ISO 13482:2014, the primary standard applicable to care robots, is not the exception. As the following paragraphs depict, there are reasons to believe that the standard fails to appropriately identify hazards and safety measures. Coupled with that, a recent technical report recognized that most of the testing methods for robotic exoskeletons have not been implemented or evaluated in terms of their accuracy and reproducibility, thus hindering developers' application of safety standards (ISO/TR 23482-1:2020 Application of ISO 13482:2014 safety-related test methods). In this sense, ISO/TR 23482-1:2020 suggests "users of this document are therefore advised to apply the tests with care."

4.1 FoF as a hazard and developing predictors for risk of falling

Fear of falling (FoF) or basophobia is among the leading causes behind reduced mobility, independence, and quality of life among elderly populations (Grimmer et al., 2020). For exoskeleton technology, the FoF can compromise and constrain the device's performance that aims to restore a regular gait pattern and not a feared one. Although such a psychological aspect is so tightly connected with safety, ISO 13482:2014 does not include FoF as a potential hazard associated with lower-limb exoskeletons. The problem is even more acute if one considers the insufficiency of studies that correlate the use of exoskeletons, self-perception, and safety.

PROPELLING proposes a method to test whether FoF is a hazard associated with lower-limb exoskeletons. It seeks to monitor different users' heart and respiration rates while using the device in different conditions (*see* figure 2). These include (i) seating, (ii) standing, and (iii) walking conditions while wearing the exoskeleton. It also includes fear-related

[4] The LIAISON project, which stands for *Liaising robot development and policymaking*, aims to ideate an alignment model between robots' legal appraisal channeling robot policy development from a hybrid top-down/bottom-up perspective to solve the mismatch between robot development and policymaking activities. LIAISON focuses specifically on personal care robots (ISO 13482:2014), rehabilitation robots (IEC 80601–2–78–2019), and agricultural robots (ISO 18497:2018). As part of this effort, the LIAISON project established a set of surveys to obtain insight from the wider community of stakeholders - comprising robot developers, policymakers, academia, and other industry experts - into the uncovered challenges experienced within the relevant policy documents.

scenarios. The second set of testing protocols entails perturbing users while wearing the device and measuring its recovery from those perturbations. Both settings are complemented with questionnaires aimed at measuring their mental stress and heart acceleration.

These tests generate data that can be measured against specific performance indicators —for example, the falls efficacy scale— that are expected to work as predictors of the presence of FoF. Furthermore, the time and facility to recover from body sway allows correlating that fear to potential safety hazards. This set of testing can also make FoF act as a predictor of actual risks of falling. Although many exoskeleton-related studies mention the risk of falls, most of them conclude that there is no risk of falls just because no falls were observed in clinical trials (He et al., 2017). However, as He and colleagues point out, "it is likely that the safeguards and task conditions followed in those clinical trials will be distinct from those imposed by settings outside the clinic" (ibid). In short, it is possible that "the risk of actual falls in these studies was completely mitigated because of the overhead harness" and other conditions of the experimentation (ibid).

The problem here, thus, is one of choosing the adequate experimentation setting. Erroneous scenarios or designs not reproducible in further research bring the risk of being incapable of informing regulators. Moreover, poor designs overlook significant risks, such that potential hazards might remain out of the radar of policymakers. Indeed, a missing step between experimentation and policymaking is the lack of indicators for determining the risk of falling as a hazard and the subsequent need to address the issue. That circumstance points not to the non-recognition of the hazard itself but to the lack of appropriate benchmarks in the emerging field of robotic exoskeletons.

There is, however, research pointing to the convenience of using gait changes and FoF as a reliable predictor for those risks of falling (Rivasi et al., 2020; Burker et al., 2020). PROPELLING taps on those studies to pilot the use of indicators on FoF, not only as a source of novel hazards but as an appropriate set of benchmarks for predicting the risks of falling.

Figure 2. PROPELLING scenarios, protocols, and performance indicators for fear of falling policy change.

Objective	Scenario	Protocol	Performance indicator
Investigate the correlation between FoF and safety	Characterization of user experience during exoskeleton-assisted walking	User-centered assessment of exoskeleton-assisted overground walking	Stress Perceptibility Acceptability Functionality Usability Falls Efficacy Scale (FES)
	Walking/ standing during pushes	Perturbed balance assessment	body_sway recovery_time Falls Efficacy Scale (FES) Stress Perceptibility Acceptability Functionality Usability

4.2 Initiating protective stops

Regulations not only serve the goal of safety. As these devices become more interconnected, ubiquitous and incorporate artificial intelligence, concerns relating to privacy, discrimination, dignity, trust, and cyber-security arise in robotics (Wynsberghe, 2013; Fosch-Villaronga, 2019).

Protective stops and graceful collapsing provide examples of how these risks intertwine with safety. ISO 13482:2014 invites manufacturers to define personal care robot halting but does not include specific guidelines on scenarios where protective stops should be initiated, whether there could be automated protective stops, nor measures to address faulty or hampered stops.

For instance, the standards leave the question of reachability open. In this sense, ISO 13482:2014 does not clarify when a button is quickly accessible to the user. It also does not set limits to whether other persons could activate it or not and under which circumstances. Think, for instance, on the device developed within the HiBSO project, where the button was first situated on the rear side of the exoskeleton (Baud et al., 2016). Or consider those circumstances, contemplated within the ISO 13482:2014, where the protective stop may be automatically initiated. These circumstances could not only hamper reachability by the user, but they could also compromise their safety if commenced at the wrong moment. They might also expose the user to vandalisms, disagreeable jokes, or cyberattacks that could lead to fatal outcomes (Fosch-Villaronga, 2017).

Furthermore, the insertion of protective stops may translate into dependable devices that cannot be used without supervision and which might lead the user to feel a loss of control as she is no longer capable of stopping the robotic device. Relatedly, no protective stop helps the user or the caregiver 'stop' the processing of personal data, nor is there a test for validating protective stop mechanisms and graceful collapsing (Fosch-Villaronga, 2019).

Figure 3. PROPELLING scenarios, protocols, and regulatory concerns for protective stops

Objective	Scenario	Protocol	Performance indicator
Determine the correlation between protective stops and instability	Ascending/ Descending Stairs	User Exoskeleton Interaction Observation	stride_time_right/left stance_time_right/left swing_time_right/left
	Ascending/ Descending slopes	Slope walking up Slope walking down	Mediolateral MoS Anteroposterior MoS LDS Left/Right Step Length & Width Foot Placement Estimates Explained Variance by a Linear Foot Placement Mode

To tackle that lack of guidance, PROPELLING suggests executing two scenarios, one replicating unstructured environments and the other characterizing impacts on muscle coordination, to specify instances in which protective stops should be automatically activated or when their activation puts greater danger to the user (*see* figure 3). Here, the used benchmarks aim at providing information on the interaction between users of different ages and genders and the exoskeleton. Those indicators inform the implementation of design measures and provide data for the standardization of testing methods. They also provide information that could correlate with other indicators to discover the impact of cyberattacks and faulty features on users' safety. This may suggest the need to standardize novel safeguarding alternatives and testing criteria for cyber-physical systems and protective stops, potentially impacting standards for personal care robots.

4.3 Instability in collision

Another critical yet misapprehended hazard is instability in a collision. Indicators about different subjects' ability to maintain balance might suggest improving the ISO 13482:2014 to address collision instability specifically, which remains a hazard that is not associated with lower-limb exoskeletons. Furthermore, information about stability in unstructured environments could indicate validating measures to address the risk of falls. PROPELLING aims to use users' stability as an indicator of the capability of different subjects to maintain balance while wearing an exoskeleton. Furthermore, data on different ages and genders suggests specific standards for distinct categories of subjects.

This scenario points to a further challenge for harnessing robot experimentation to optimize regulatory frameworks. In light of the lack of standardized methods, there is a risk that the evidence gathered will fail to frame the technology appropriately. Without the tools for unearthing safety hazards in a manner that could be replicable, we run the risk of ending up with off-shot experiments that lack the necessary reproducibility for informing general regulations (*see* figure 4).

Figure 4. PROPELLING scenarios, protocols, and regulatory concerns for instability in collision

Objective	Scenario	Protocol	Performance indicator
Determine the correlation between the use of exoskeletons and increased instability in irregular surfaces	Walking through irregular terrains	Walking through irregular terrains	margin_of_stability gait_parameters_var ia bility

Hence, PROPELLING aims at unearthing information on whether gaps in test facilities, particularly concerning the determination of risks of collision and instability, fail to provide adequate data to inform regulators. This information is essential to understand what gaps the testing facility has in providing a good variety of scenarios to test different safety requirements (for instance, on slippery terrains).

5 Towards evidence-based policymaking: early lessons learned and discussion

Certainly, evidence tells us "what works" and "what are the risks.", especially within the context of the development of new technologies. Still, one can continue by asking what kind of evidence will deliver the necessary results, how reliable the information should be, or which sources deserve attention. These questions point to an unresolved issue in the regulation of emerging technologies, and the scenario is very much prone to cherry-picking, obfuscation, or manipulation of pieces of evidence, with the risk of regulatory failure on the horizon. If we claim to base regulation on evidence, we first need to discuss and thoroughly understand what counts as evidence to this end. Otherwise, even the best policy proposals will fail to deliver adequate results in practice.

An evidence-based approach is as good as the type of evidence gathered; the quality of that evidence depends on the quality and reproducibility of the method used and its collection. Still, the field of robotics largely lacks such methodologies. The problem with lower-limb exoskeletons and robotics, in general, is the paucity of agreed methods for evaluating certain parts of such developments, mainly the psychological, but also other aspects such as trust, dignity, and privacy. This lack has been confirmed in a survey of the state-of-art of exoskeletons conducted by ISO. In the technical report, it was acknowledged that the test methods for exoskeletons are not put into action or broadly evaluated regarding their accuracy and reproducibility. Also, ISO 13482:2014 stated that the Working Group would revise the definition of personal care robots, considering concrete users such as children, elderly persons, and pregnant women, but the revised standard in 2020 nevertheless shows no changes in this respect (Fosch-Villaronga & Drukarch, 2021).

The second missing step between experimentation and policy-making is the lack of indicators for determining the existence of a hazard and the subsequent need for addressing the issue. Identifying what aspects constitute a safety hazard and which aspects should be evaluated and safeguarded are extremely difficult to reach without field knowledge. On top of that, there is nothing like the medical ideal of serving a patient's interests in technology regulation. In robotics, for example, ISO 13482:2014 does not define personal care. It has many confusing categories that seem to suggest avoiding compliance with the medical device regulation rather than ensuring users' safety (Fosch-Villaronga, 2016). However, regulations not only serve the goal of safety, and if they do, safety has many dimensions, not only physical (Martinetti et al., 2021). As these devices become more interconnected, ubiquitous and incorporate artificial intelligence, other concerns of privacy, discrimination, and even cyber-security arise. The findings published by the LIAISON project highlight this issue and stress that although each of the standards investigated within the context of the project, including ISO 13482:2014, is concerned with physical safety requirements, the legislative system includes many other fundamental rights to be protected. In this line of thought, insights from the broader community of stakeholders within the context of this project have indicated that standards should shift from mono-impact to multi-impact. In this sense, they should include factors related to ethics, environmental sustainability, liability, accountability, privacy, data protection, and psychological aspects, further stressing the need for a multi-disciplinary multi-stakeholder approach in crafting policies aimed at framing new technologies.

These findings shed light on the fact that safety standards may not be offering enough protection to users and also points to the lack of benchmarking in the emerging field of robotics. Moreover, agreement on a particular type of test is not sufficient. It is also crucial to identify appropriate indicators for determining what counts as a relevant hazard and what pieces of evidence serve to identify it. Otherwise, experimentation results cannot be generalized, and developers will keep recurring to different experiments with varied results. That is the case, for instance, with the strategies most marketed exoskeletons employ for dealing with potential falls. Here, the lack of a settled methodology has led to several strategies whose effectiveness and risks remain unclear (He et al., 2017).

Moreover, a proposal for harnessing experimentation should also be mindful of regulatory bottlenecks. Experimentation methods should not only be appropriate for unearthing needs and deficiencies. For institutions and instruments to keep an adequate connection with the stream of technological innovation, these methods should also suit the procedures and practices behind regulatory development (Brownsword, 2017), which can generally be divided into public and private policymaking initiatives.

Public policymaking can be generally defined as a system of laws, regulatory measures, courses of action, and funding priorities concerning a given topic promulgated by a governmental entity or its representatives. There are examples of public policymaking activities within the field of new technologies at the European level, resulting in a wide variety of EU-wide measures, particularly Directives and Regulations. The EU nevertheless struggles to release technology-savvy, sector-specific guidelines (Fosch-Villaronga, 2019). Public policymaking is often outdated and tech-neutral, and legal responsiveness does not always follow technological development timely or at all as a consequent step (Collingridge, 1980; Marchant, 2011; Newlands et al., 2020).

To address this inability of regulators to keep up with the fast pace of innovation and propose regulatory actions matching state of the art and the foreseeable impacts such emerging technologies may have, private actors have developed private standards, such as the ISO standards, to mitigate the ethical and legal risks and concerns posed by robotics. ISO standards are developed within national groups involving different levels of access and influence on the system. Furthermore, they involve experts from all over the world gathered within technical committees that join together consumer associations, academia, NGOs, and government. They allow technology to work seamlessly and establish trust so that markets can operate smoothly, thereby providing a common language to measure and evaluate performance; b) make interoperability of components made by different companies possible, and c) protect consumers by ensuring safety and durability and market equity. The development of ISO standards involves a consensus-based approach, which means that proposals are submitted to repeated consideration until all the parties with voting prerogatives agree on the definitive form (cf. Fosch-Villaronga and Golia, 2019).

Although these soft-law instruments are excellent for reaching international agreements on relevant topics, they have also raised practical concerns. Standardization generally involves a multiplicity of actors bringing in various complexities and shifts the centralization of regulation from public democratic processes to private ones that do not guarantee the rule of law. Furthermore, standards in themselves are not binding (if not included in contractual clauses); they do not fix consequences for violations or noncompliance, are usually mono-impact, that is, they focus on one aspect (e.g., on safety, often leading to a multiplicity of standards for one particular domain), and generally include confusing categories (Fosch-Villaronga, 2016).

Early stages of empirical research confirm these concerns. Van Rompaey et al. (2021) presented evidence that what often prevails among roboticists are economic and safety concerns, whereas they often feel pressed to rely on a set of tests that they deem to be insufficient. Remarkably, the authors pointed to developers' overconfidence in the safety of their products, with most of them relying on the device's capability to enter in a safe mode or to shut off as the sole safeguard (ibid).

In light of these difficulties, two lessons must be posited at the early stages of PROPELLING. First, testbeds represent an experimental, co-creative approach to innovation policy that aims to test, demonstrate, and advance new socio-technical arrangements under real-world conditions. In the simplest sense, thus, testbeds are controlled experimental spaces that facilitate a kind of performance or hypothesis testing under presumably realistic conditions (Engels et al., 2019.). Hence, they set a promising environment for developing generalizable benchmarks and testing methods for robotics and emerging technologies. They provide an adequate landscape for moving towards agreed yardsticks and experiments. In this sense, testbeds are data generators for better policymaking and set the stage for discussing adequate experiments, methods, and performance indicators for appraising safety in a manner that can be replicable across domains.

The second lesson is the convenience of creating Shared Data Repositories (SDRs) and connecting them to policymaking. The aim is to gather data concerning compliance with the law of a particular project, robot use, or development and use it as evidence for policymaking (Fosch-Villaronga & Heldeweg, 2018; Fosch-Villaronga and Golia, 2018). These SDRs could take the form of databases compiling the results of reproducible experiments and risk assessments, along with related robot legislation/regulation collected over time and across many projects (ibid). Documenting and formalizing these processes (in lessons learned) would allow the regulatory framework to have a grounded knowledge and understanding of what characteristics and regulatory needs new robot technologies have, and this knowledge could be highly valuable for future developers and policymakers.

Generating and sharing this information would help forge a community revolving around the development of robotics that also would consider the policymaking side. For instance, consider the LIAISON project. LIAISON envisions an iterative regulatory process for robot governance, a theoretical model that represents a practical step forward in coordinating and aligning robot and regulatory development, called the Iterative Learning Governance Process (ILGP) (Fosch-Villaronga & Heldeweg, 2019). The primary outcome of the LIAISON project is to develop the design concept for liaising robot development and policymaking to increase overall robot safety. This design concept will further develop the Iterative Regulatory Process for Robot Governance, which was ideated as a theoretical model that links technology impact assessments to legislative ex-post evaluations via shared data repositories intending to create evidence-based policies that can serve as temporary benchmarks for future and new uses or robot developments (Fosch-Villaronga & Heldeweg, 2018; 2019). In other words, the knowledge extracted from these assessments that would match robots and legislative affordances and limitations could be collected and saved in an SDR, allowing for the creation of an interactive learning process.

In this vein, developing replicable experiments and agreed yardsticks and complementing them with data repositories accessible to regulators is the path forward to evidence-based standard making. Linking experimentation settings with standard-making processes can speed the creation, revision, or discontinuation of norms governing robot technology, increase their effectiveness in ensuring overall safety, and the legal certainty regarding a fast-paced changing environment like robotics. This process can shed light on what needs regulatory attention for adequate robot governance and make robot oversight more precise and concrete, allowing for easier compliance and virtually not wasting the potential testing zones to generate relevant knowledge for policymaking.

6 Conclusion

Gathering information is vital for framing technology adequately, and experimentation serves as an appealing source of data on emerging developments for this purpose. However, harnessing that information largely depends on whether that evidence comes from replicable and generalizable methods that fit the regulatory process. Research on those aspects in the field of regulatory interventions for robotics and modern technologies currently remains in its infancy.

This paper introduced PROPELLING as a stepping stone in ideating practical ways to use benchmarking ecosystems such as EUROBENCH as data generators for policymaking. This process can make robot oversight more precise and concrete, allowing for easier compliance and virtually not wasting the potential testing zones to generate relevant knowledge for policymaking.

Igniting a conversation on the limitations and promises of evidence-based regulation is crucial for steering EU regulatory discussions to more practical and definite regulatory reform proposals. In this vein, proposals will become responsive to robot development while moving away from mere ethical recommendations that further confuse the already very complex robot legal panorama (HLEG Trustworthy AI Assessment List, 2019).

Acknowledgments

This paper is part of a project that has received funding from the European Union's Horizon 2020 research and innovation program, via an Open Call issued and executed under Project EUROBENCH (grant agreement No. 779963).

References

Alemzadeh, H., Raman, J., Leveson, N., Kalbarczyk, Z., and Iyer, R. K. (2016). Adverse events in robotic surgery: a retrospective study of 14 years of FDA data. PloS one, 11(4), e0151470.

Aymerich-Franch, L., and Ferrer, I. (2020). The implementation of social robots during the COVID-19 pandemic. arXiv preprint arXiv:2007.03941.

Bardaro, G. El-Shamouly, M., Fontana, G., Awad, R., and Matteucci, M. (2019). Toward model-based benchmarking of robot components- IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019, pp. 1682-1687.

Baud, R., Ortlieb, A., Olivier, J., Bouri, M., and Bleuler, H. (2017). HiBSO Hip Exoskeleton: Toward a Wearable and Autonomous Design. New Trends in Medical and Service Robots, 48, 185-195.

Bergamasco, M., and Herr, H. (2016). Human–robot augmentation. In B. Siciliano & O. Khatib (Eds.), Springer handbook of robotics (2nd ed., pp. 1875–1906). Cham: Springer.

Brownsword, R. (2008). Rights, Regulation, and the Technological Revolution. Oxford: Oxford University Press.

Brownsword, R., and Goodwin, M. (2012). Regulatory connection II: Disconnection and sustainability. In Law and the Technologies of the Twenty-First Century: Text and Materials (Law in Context, pp. 398-420). Cambridge: Cambridge University Press.

Brownsword, R. (2017). Law, Liberty, and Technology. In The Oxford Handbook of Law, Regulation and Technology (1st ed., Oxford Handbooks, pp. The Oxford Handbook of Law, Regulation and Technology, 2017-07-20). Oxford University Press.

Carr, N. (2010). The shallows: How the internet is changing the way we think, read and remember. Atlantic Books Ltd.

Collingridge D. (1980) The Social Control of Technology. St. Martin's Press, New York.

Engels, F., Wentland, A., and Pfotenhauer, S.M. (2019). Testing future societies? Developing a framework for test beds and living labs as instruments of innovation governance. Research Policy, 48(9), 103826.

Fosch-Villaronga, E. (2016). ISO 13482:2014 and Its Confusing Categories. Building a Bridge Between Law and Robotics. In Wenger P., Chevallereau C., Pisla D., Bleuler H., Rodić A. (eds) New *Trends in Medical and Service Robots*, Vol. 39, Series Mechanisms and Machine Science, Springer, 31-44. https://doi.org/10.1007/978-3-319-30674-2_3.

Fosch Villaronga, E. (2017). Towards a legal and ethical framework for personal care robots: analysis of person carrier, physical assistant and mobile servant robots/per Eduard Fosch Villaronga; director: Dr. Antoni Roig Batalla; co-director: Dr. Jordi Albiol Canals.

Fosch Villaronga, E. and Heldeweg, M. (2018). 'Regulation, I Presume?', Said the Robot. Towards an Iterative Regulatory Process for Robot Governance. SSRN Electronic Journal. 10.2139/ssrn.3194497.

Fosch-Villaronga, E., Felzmann, H., Pierce, R. L., De Conca, S., De Groot, A., Ponce Del Castillo, A., and Robbins, S. (2018). 'Nothing comes between my robot and me': Privacy and human-robot interaction in robotised healthcare. In: R. Leenes, R. van Brakel, S. Gutwirth, and P. De Hert (Eds.), Data protection and privacy: The internet of bodies.

Fosch Villaronga, E., and Golia, A. (2019). Robots, standards and the law: Rivalries between private standards and public policymaking for robot governance. The Computer Law and Security Report, 35(2), 129-144.

Fosch-Villaronga, E. (2019). Robots, healthcare, and the law: Regulating automation in personal care. Routledge.

Fosch-Villaronga, E., & Drukarch, H. (2021). On Healthcare Robots: Concepts, definitions, and considerations for healthcare robot governance. *arXiv preprint arXiv:2106.03468*

Fosch-Villaronga, E., & Mahler, T. (2021). Cybersecurity, safety and robots: Strengthening the link between cybersecurity and safety in the context of care robots. *Computer Law & Security Review*, *41*, 105528.

Fosch Villaronga E. & Drukarch H. (2021), H2020 COVR FSTP LIAISON – D1.2 Report on usefulness of LIAISON no. MS1. Leiden: eLaw / Leiden University.

Grand View Research. (2020). Exoskeleton market size worth $4.2 billion by 2027|CAAGR: 2266.33%.
San Francisco. https://www.grandviewresearch.com/press-release/global-exoskeleton-market.

Guihot, M., and Moses, L. (2020). Artificial intelligence, robots and the law. Australia: LexisNexis.

Greenbaum, D. (2015a). Exoskeleton progress yields slippery slope. Science, 350(6265), 1176. https://doi.org/10.1126/science.350.6265.1176-a.

Greenbaum, D. (2015b). Ethical, legal and social concerns relating to exoskeletons. ACM SIGCAS Computers and Society, 45(3), 234–239.

Grimmer, M., Riener, R., Walsh, C.J., and Seyfarth, A. (2019). Mobility related physical and functional losses due to aging and disease - A motivation for lower limb exoskeletons. Journal of Neuroengineering and Rehabilitation, 16(1), 2.

He, Y., Eguren, D.L., Trieu P., and Contreras-Vidal, J.L. (2017). Risk management and regulations for lower limb medical exoskeletons: A review. Medical Devices (Auckland, N.Z.), 10, 89-107.

High-Level Expert Group (HLEG) on Artificial Intelligence, HLEG AI (2019) Ethical Guidelines for Trustworthy AI. European Commission. Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (last accessed 12 March 2021).

International Organization for Standardization. (2014). Robots and robotic devices—Safety requirements for personal care robots (ISO 13482:2014). https ://www.iso.org/stand ard/53820 .html.

International Organization for Standardization. (2020).Robotics — Application of ISO 13482 — Part 1: Safety-related test methods (ISO/TR 23482-1:2020). https://www.iso.org/obp/ui/#iso:std:iso:tr:23482:-1:ed-1:v1:en.

Kapeller, A., Felzmann, H., Fosch-Villaronga, E., & Hughes, A. M. (2020). A taxonomy of ethical, legal and social implications of wearable robots: an expert perspective. Science and Engineering Ethics, 1-19.

Kapeller, A., Felzmann, H., Fosch-Villaronga, E., Nizamis, K. Hughes, A. M. (2021) Implementing Ethical, Legal, and Societal Considerations in Wearable Robot Design. *Applied Sciences*, *11*(15), 6705 https://www.mdpi.com/2076-3417/11/15/6705.

Koops, B., Ten Dimensions of Technology Regulation - Finding Your Bearings in the Research Space of an Emerging Discipline (2010). Dimensions of Technology Regulation, pp. 309-324, M.E.A. Goodwin et al., eds., Nijmegen: WLP, 2010 , Tilburg Law School Research Paper No. 015/2010.

Leenes, R., Palmerini, E., Koops, B., Bertolini, A., Salvini, P. and Lucivero, F. (2017). Regulatory challenges of robotics: some guidelines for addressing legal and ethical issues. Law, Innovation and Technology, 9:1, 1-44.

Leigh, A. (2009). Evidence-Based Policy: Summon the Randomistas?. Paper delivered to a Productivity Commission Roundtable on 'Strengthening Evidence-Based Policy in the Australian Federation'.

Lipton, Z. (2018). The mythos of model interpretability. Communications of the ACM, 61(10), 36-43.

Mann, S. (2012). Wearable computing. In M. Soegaard and R. F. Dam (Eds.), The encyclopedia of human– –Computer interaction (2nd ed.). Interaction Design Foundation. https://www.interaction-desig.

Marchant, G.E., Allenby, B.R. and Herkert, J.R. eds., (2011). The growing gap between emerging technologies and legal-ethical oversight: The pacing problem (Vol. 7). Springer Science & Business Media.

Martinetti, A., Chemweno, P., Nizamis, K., & Fosch-Villaronga, E. (2021) Redefining safety in light of human-robot interaction: a critical review of current standards and regulations. *Frontiers in Chemical Engineering*, https://www.frontiersin.org/articles/10.3389/fceng.2021.666237/abstract.

Morante, S., Victores, J. G., and Balaguer, C. (2015). Cryptobotics: Why robots need cyber safety. Frontiers in Robotics and AI.

Newlands, G., Lutz, C., Tamò-Larrieux, A., Fosch-Villaronga, E., Scheitlin, G., and Harasgama, R. (2020). Innovation under Pressure: Implications for Data Privacy during the Covid-19 Pandemic. Big Data & Society, SAGE, 7(2), 1-14.

Palmerini, E., Bertolini, A., Battaglia, F., Koops, B. J., Carnevale, A., and Salvini, P. (2016). RoboLaw: Towards a European framework for robotics regulation. Robotics and autonomous systems, 86, 78-85.

Rocha, C., and Harris, J. (2019). Evidence-based policymaking in the food–health Nexus. IDS Bulletin (Brighton. 1984), 50(2), 57-72.

Sandel, M. (2007). The case against perfection : Ethics in the age of genetic engineering. Cambridge, Mass: Belknap Press of Harvard University Press.

Singh, H., Meyer, A. N., & Thomas, E. J. (2014). The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations. BMJ quality & safety, 23(9), 727-731.

Vallor, S. (2011). Carebots and caregivers: Sustaining the ethical ideal of care in the twenty-first century. Philosophy & Technology, 24(3), 251-268.

Van Rompaey, L., Jønsson, R., Jørgesen, K.E. (2021) Designing lawful machine behaviour. Roboticists' legal concerns. [Conference presentation]. EURA Conference 2021 - Regulating UncertAInty.
https://www.youtube.com/watch?v=GjCMRu86IRU&t=9461s.

Valori, M., Scibilia, A., Fassi, I., Saenz, J., Behrens, R., Herbster, S. Bidard, C., Lucet, E., Magisson, A., Schaake, L., Bessler, J., Prange, G., Kühnrich, M., Lassen, A., and Nielsen, K. (2021). Validating Safety in Human–Robot Collaboration: Standards and New Perspectives. Robotics (Basel), 10(2), 65.

Weng, Y. H., Sugahara, Y., Hashimoto, K., & Takanishi, A. (2015). Intersection of "Tokku" special zone, robots, and the law: a case study on legal impacts to humanoid robots. *International Journal of Social Robotics*, *7*(5), 841-857.

Wynsberghe, A. van (2013). Designing robots for care: Care centered value-sensitive design. Science and engineering ethics, 19(2), 407-433.