Safety of the Intended Functionality: What about Mental Harm?

Barbara Gallina
Mälardalen University, Västerås, Sweden, Email: barbara.gallina@mdu.se

***Abstract*—Safety Of The Intended Functionality (SOTIF) is defined in ISO 21448:2022 as absence of unreasonable risk due to hazards resulting from functional insufficiencies of the intended functionality or its implementation. The risk is defined as combination of the probability of occurrence of harm and the severity of that harm. Harm is typically intended as physical harm. Recent studies however have pointed out that with the increase of the automation level, mental conditions might be worsened or emerge. Hence, we state that mental harm shall be taken into consideration to avoid violating a basic human right (the right to mental health). Hence, in this abstract, we aim at proposing an extended interpretation of the notion of harm while conducting risk assessment in the SOTIF context.**

***Keywords*—SOTIF, Mental Harm, Severity, Controllability.**

I. Introduction and background

ISO 21448:2022 [1] defines (definition 3.25) Safety Of The Intended Functionality (SOTIF) as absence of unreasonable risk due to hazards resulting from functional insufficiencies of the intended functionality or its implementation. The risk is defined as combination of the probability of occurrence of harm and the severity of that harm. While conducting the SOTIF-HARA (Hazards Analysis and Risk Assessment) the severity of harm, and the controllability of hazardous events, can be estimated using the method described in ISO 26262-3:2018 [2], Clause 6. In ISO 26262, harm is defined as physical injury or damage to the health of persons. Hence, this definition does not exclude mental harm. As far as we know, however, the focus has so far been on physical harm. As a consequence, harm is interpreted in a narrow manner. Specifically, the Abbreviated Injury Scale (AIS) is used for estimating the severity. The estimation of the controllability does not seem to take into consideration the mental conditions that may manifest themselves as a consequence of the increased automation level. ISO 21448:2022 highlights that the observed outcome and the estimated parameters for a specific hazard can be different for the SOTIF-HARA. However, the focus has so far been on physical harm. Recent studies (e.g., [3]) have pointed out that with the increase of the automation level, mental conditions (such as new situational phobias like automato-phobia) might emerge. Hence, we propose to consider mental harm when estimating the severity and controllability. The rest of the abstract is organised as follows. In Section II, we present our proposal. In Section III, we conclude and sketch future work.

II. What about mental harm?

In this abstract, to raise the quality bar while conducting the SOTIF-HARA, we propose to extend the interpretation of harm. This would contribute to: 1) increasing the chances of better estimating the controllability of the driver (which are the underlying assumptions on the mental health of the driver?) in e.g., regaining the control of the driving task at SAE J3016 level 3 as well as other road users; 2) deepening the brainstorming regarding direct misuse. It is well known that drivers with psychiatric disorders (such as mild to moderate anxiety or depression) may drive. However, it is unknown if increased automation may worsen their mental conditions and hence increase the risk for misuse; 3) reaching a better understanding of the balance while using the GAMAB (Globalement Au Moins Aussi Bon) risk acceptance principle since the consideration would not be limited to the physical harm (which in a long-term future might decrease) but would also consider the mental harm, which instead in a long-term future might increase, if not-considered. To the best of our knowledge, our work represents a novelty in its proposal for expanding the interpretation of harm in the automotive SOTIF context. This work, conducted within the 4DSafeOps [4] project, cross-fertilises the automotive domain with what was proposed in [5], where the need of considering mental harm in the medical domain was emphasised.

III. Conclusion and Future Work

To improve SOTIF, we proposed to expand the interpretation of harm while conducting the HARA. Our proposal could also fit ISO 26262-HARA. As future work, we aim at conducting a systematic literature review in relation to the mental conditions that emerged or were proven to be worsened and their quantifiability and coverage by tort law. This with the purpose of providing evidence for a future revision of ISO 21448:2022, in addition to what proposed in [6].

Acknowledgment: We thank the anonymous reviewers.

References

[1] *International Organization for Standardization (ISO)*, "ISO 21448: Road vehicles — Safety of the intended functionality (SOTIF)," 2022.

[2] International Organization for Standardization (ISO), "*ISO 26262:2018 - Road vehicles – Functional safety*," 2018.

[3] G. Meinlschmidt, E. Stalujanis, L. Grisar, M. Borrmann, and M. Tegethoff, "Anticipated fear and anxiety of automated driving systems: Estimating the prevalence in a national representative survey," *Int. J. Clin. Health Psychol. (IJCHP)*, vol. 23, no. 3, p. 100371, 2023.

[4] 4DSafeOps Team, "4DSafeOps, Standards-Assurance Case-Process-Product-Aware SafeOps #49, Software Center." [Online]. Available: https://www.software-center.se

[5] J. L. de La Vara, B. Gallina, A. Fernández-Caballero, J. P. Molina, A. S. García, and C. Ayora, "Assurance of software-intensive medical devices: What about mental harm?" in *53rd IEEE/IFIP Int. Conf. on Dependable Systems and Networks - Supplemental Vol. (DSN-S)*, 2023, pp. 168–172.

[6] O. Manabu and B. Gallina, "Safety of the intended functionality of external human interfaces: Gaps and research agenda," in *48th IEEE International Conference on Computers, Software, and Application (COMPSAC)*, 2024.