Name: Marco Sanchi

Position: PhD Student-Researcher

Academic Affiliation: National Italian PhD on AI & Society, University of Bologna - University of Pisa

Presentation Title: "Artificial Intelligence War Crimes: Regulating Accountability".

Summary and Context:

This presentation explores the issue of allocating personal liability for war crimes committed by autonomous Artificial Intelligence (AI) systems, advocating for innovative accountability models to address a crisis of causality present in the domain.

Humanitarian Law and International Criminal Law recognize war crimes as serious violations that substantially breach the former and formally violate the latter. Decades of jurisprudence from the International Criminal Court, and scholarly debate have established a necessary condition for punishing war crimes: responsibility for those who perpetrated the *actus reus*.

In modern warfare, this standpoint is challenged by autonomous AI weaponry: systems capable of executing military actions without direct human control. These AI-powered weapons have seen a significant surge in research, development, and production, dominating military spending and funding and becoming more active in contemporary theaters of war.

Recognizing the increasing role of AI-powered arms in the future of warfare, it is paramount to govern and regulate the risks, challenges, and unwarranted consequences such systems may bring about, starting with allocating accountability for unplanned or unforeseen war crimes perpetrated by such machines.

Traditional criminal liability models fail to establish personal accountability. They exhibit a fundamental inability to address this phenomenon adequately, contributing to the crisis of causality debated by criminal law scholars. As a matter of fact, the standard causality chain and chain of command fall short of justifying the personal scapegoating of a military officer in the aftermath of unplanned war crimes committed autonomously by defective AI-powered arms.

This raises numerous legal and ethical concerns.

To bridge this gap, the presentation introduces a liability model centered on administrative liability for third-party criminal conduct, drawing inspiration from corporate liability schemes. The presentation concludes by questioning whether it is opportune to direct this model toward holding military personnel accountable or making developers of AI-powered arms *prima facie* liable.

Above all, the presentation identifies a gap in the current liability model for charging and punishing war crimes committed by autonomous AI systems. The problem is framed as crucial for the military domain, and a solution focused on regulating it through administrative liability is presented. An open question is introduced to spark debate and discussion, centered on which stakeholders should be potentially held accountable.